SCIENCE LIBRARY

Vol. XV, No. 5 PSYCHOLOGICAL REVIEW PUBLICATIONS OCTOBER, 1932

OCT 10 1932

Journal of Experimental Psychology

EDITED BY
SAMUEL W. FERNBERGER, UNIVERSITY OF PENNSYLVANIA

HOWARD C. WARREN, PRINCETON UNIVERSITY (*Review*)
WALTER S. HUNTER, CLARK UNIVERSITY (*Index*)
HERBERT S. LANGFELD, PRINCETON UNIVERSITY (*Monographs*)
EDWARD S. ROBINSON, YALE UNIVERSITY (*Bulletin*)

CONTENTS

The Influence of Punishment during Learning upon Retroactive Inhibition: M. E. BUNCH AND F. D. MCTEER 473
Transfer of Response in Human Subjects: H. GULLIKSEN 496
Learning during Directed Attention: W. C. F. KRUEGER 517
A Study of the Eye-Movements of Stutterers during Oral Reading: H. H. JASPER AND E. MURRAY 528
Habituation, Retention and Perseveration Characteristics of Direct Waking Suggestion: E. F. PATTEN, ST. C. A. SWITZER AND C. L. HULL 539
Is the Judgment of the Group Better than that of the Average Member of the Group?: J. R. STROOP 550
An Experimental Study of Vividness in Learning and Retention: W. L. VAN BUSKIRK 563
Chromatic Sensations from Field-Size Measurements at Varying Wavelength: C. A. RINDE 574
Effect of Background on Visual Acuity of Circle Grids: E. B. GREENE 585
A New Type of Psychogalvanometer: T. H. HOWELLS 591
A New Device for the Measurement of Time Intervals: F. M. DENTON 598

PUBLISHED BI-MONTHLY
FOR THE AMERICAN PSYCHOLOGICAL ASSOCIATION
BY THE PSYCHOLOGICAL REVIEW COMPANY
PRINCE AND LEMON STS., LANCASTER, PA.
AND PRINCETON, N. J.

Entered as second-class matter, July 1, 1920, at the post-office at Lancaster, Pa.

PUBLICATIONS

OF THE

AMERICAN PSYCHOLOGICAL ASSOCIATION

EDITED BY

HOWARD C. WARREN, PRINCETON UNIVERSITY (*Review*)
S. W. FERNBERGER, UNIVERSITY OF PENNSYLVANIA (*J. Exper. Psych.*)
WALTER S. HUNTER, CLARK UNIVERSITY (*Index and Abstracts*)
HENRY T. MOORE, SKIDMORE COLLEGE (*J. Abn. and Soc. Psychol.*)
HERBERT S. LANGFELD, PRINCETON UNIVERSITY (*Monographs*)
EDWARD S. ROBINSON, YALE UNIVERSITY (*Bulletin*)

HERBERT S. LANGFELD, Business Editor

PSYCHOLOGICAL REVIEW

containing original contributions only, appears bi-monthly, January, March, May, July, September, and November, the six numbers comprising a volume of about 540 pages.

PSYCHOLOGICAL BULLETIN

containing critical reviews of books and articles, psychological news and notes, university notices, and announcements, appears monthly (10 numbers), the annual volume comprising about 720 pages. Special issues of the BULLETIN consist of general reviews of recent work in some department of psychology.

JOURNAL OF EXPERIMENTAL PSYCHOLOGY

containing original contributions of an experimental character, appears bi-monthly, February, April, June, August, October, and December, the six numbers comprising a volume of about 700 pages (from Jan. 1, 1932).

PSYCHOLOGICAL INDEX

is a compendious bibliography of books, monographs, and articles upon psychological and cognate topics that have appeared during the year. The INDEX is issued annually in June, and may be subscribed for in connection with the periodicals above, or purchased separately.

PSYCHOLOGICAL ABSTRACTS

appears monthly, the twelve numbers and an index supplement making a volume of about 700 pages. The journal is devoted to the publication of non-critical abstracts of the world's literature in psychology and closely related subjects.

PSYCHOLOGICAL MONOGRAPHS

consists of longer researches or treatises or collections of laboratory studies which it is important to publish promptly and as units. The price of single numbers varies according to their size. The MONOGRAPHS appear at irregular intervals and are gathered into volumes of about 500 pages.

JOURNAL OF ABNORMAL AND SOCIAL PSYCHOLOGY

appears quarterly, April, July, October, January, the four numbers comprising a volume of 448 pages. The journal contains original contributions in the field of abnormal and social psychology, reviews, notes and news.

ANNUAL SUBSCRIPTION RATES

Review: \$5.50 (Foreign, \$5.75). **Index:** \$4.00 per volume.
Journal: \$7.00 (Foreign, \$7.25). **Monographs:** \$6.00 per volume (Foreign, \$6.30).
Bulletin: \$6.00 (Foreign, \$6.25). **Abstracts:** \$6.00 (Foreign, \$6.25).
Abnormal and Social: \$5.00 (Foreign, \$5.25). Single copies \$1.50.
Current numbers: **Journal,** \$1.25; **Review,** \$1.00; **Abstracts,** 75c; **Bulletin,** 60c.

COMBINATION RATES (from Jan. 1, 1932)

Review and Bulletin: \$10.00 (Foreign, \$10.50).
Review and J. Exp.: \$11.00 (Foreign, \$11.50).
Bulletin and J. Exp.: \$12.00 (Foreign, \$12.50).
Review, Bulletin, and J. Exp.: \$16.00 (Foreign, \$16.75).
Review, Bulletin, J. Exp., and Index: \$19.00 (Foreign, \$19.75).

Subscriptions, orders, and business communications should be sent to the

PSYCHOLOGICAL REVIEW COMPANY

PRINCETON, NEW JERSEY

Journal of
Experimental Psychology

VOL. XV, No. 5 OCTOBER, 1932

THE INFLUENCE OF PUNISHMENT DURING LEARNING UPON RETROACTIVE INHIBITION [1]

BY MARION E. BUNCH
Washington University

AND

FRANCES DAVIS McTEER
Colleges of the City of Detroit

INTRODUCTION

That the extent to which any previously acquired activity can be revived after an interval of time is dependent upon the kind and amount of activity which has taken place in the interval between learning and reinstatement is a generally accepted psychological phenomenon. In those instances where the intervening activity interferes with the reinstatement of the previously acquired activity, the phenomenon has frequently been termed *retroactive inhibition* despite criticism of this term as implying an illogical explanation of the phenomenon. Following McGeoch's (4) defense of the term on the basis of its widespread usage, it seems justifiable to use the term here merely as the name of a psychologically well-known phenomenon and without any reference to an explanatory concept.

Assuming that the efficiency with which the previously acquired activity can be reinstated is a measure of retention, and, assuming further that the degree of retention is in part a

[1] The experimental data upon which this report is based were collected at Washington University, St. Louis, during 1927–28.

function of the conditions under which the original learning occurred, then it is probable that any condition which markedly influences the original learning may also influence its retention, and therefore would influence the degree of retroactive inhibition when the circumstances are such that this phenomenon is likely to be present. By a similar inference, conditions which markedly affect the interpolated activity may also be expected to have some influence upon the degree of the retroactive effect.

Evidence in support of such inductions may be derived from several experiments already done in this field. Tolman (**6**) after experimenting with several materials and under a variety of conditions, concludes that under more stimulating conditions fewer repetitions are required for learning, and that the material so learned is less susceptible to interference by interpolated learning. The distinctions between the various stimulating conditions were made in terms of the efficiency of the original learning and of the interpolated learning, namely, pleasant words as compared with indifferent words, normal learning with learning under the distraction of tapping, normal learning with learning after taking caffeine, and efficient with inefficient working periods (morning and evening). Skaggs (**5**) repeated Tolman's condition of morning and evening learning and compared the results thus obtained with those where the first and second halves of the day's experiment were compared; he found that interference was greater when the subject was tired irrespective of the time of the day. Whitely (**8**) presents data which indicate that the efficiency of retention is influenced by the set with which the subjects approach the original learning. In this experiment he used a situation which he termed 'the arousal of an apperceptive set,' that is, the presentation of an association test in a particular field such as to bring into action whatever associative systems may belong to that field. This aroused set is 'congruous' when it is in the same field as the material to be learned. The arousal of an apperceptive set immediately preceding learning has no influence on that learning but has a detrimental effect upon recall 24 or 48 hours afterward;

the interference is greater when the set is congruous. Harden (**3**) has studied the effect of 'emotional reactions' on retention and found that recall of nonsense syllables was less efficient when, during the two minute interval between learning and recall, the subject's oral reading was distracted by such events as, three to six rather severe electric shocks on both wrists, the collapse of the chair in which the subject was sitting, or, the firing of a revolver within a few feet of the subject. If this oral reading be regarded as the interpolated activity, then Harden's results support the assumption that the conditions under which the interpolated activity takes place have some influence upon the capacity of that interpolation to interfere with a previously acquired activity.

These experiments all lend support in varying measure to the above assumptions that any conditions which markedly influence the original learning or the interpolated work are very likely to have some influence upon the degree of retroactive inhibition. It is the purpose of the present investigation to make a further study in this field by employing a variation in the conditions of motivation during learning.

In an earlier study by the senior author (**1**) it was found that with human adults electric shock as punishment for errors in maze learning increased remarkably the subjects' learning efficiency as measured by the number of errors, the number of trials, and the total amount of time required to master the problem.

It had also been previously demonstrated that maze habits in human adults were susceptible to interference when another maze situation was interpolated between the learning of the original solution and its recall (**7**).

In view of these known results, human maze learning was chosen as the problem and the administering of electric shock in punishment for errors as the experimental condition whose influence upon retroactive inhibition was to be tested.

Thus, the problems of this study are:

1. What is the influence of administering electric shock in punishment for certain errors during the learning of a stylus maze upon the susceptibility of this problem to interference

from the acquisition of a second maze solution in the interval before the relearning of the first?

2. What is the influence of administering punishment for certain errors during the learning of the interpolated maze solution upon the detrimental effect which the latter will have upon the retention of a previously learned maze problem?

Conditions of the Experiment

Apparatus.—Two stylus mazes comprised the material for learning and interpolation. Both mazes were of such a character that the subject could be given an electric shock when the metal stylus came in contact with the end of a blind alley. Maze 1, that used as the original learning problem for all subjects, was made of brass and of the same pattern as the one employed by Carr in his study of the influence of visual guidance in maze learning (2). The construction of this maze has been described in detail in the report of the previous investigation with punishment for maze learning (1).

The second maze, the interpolated problem, was constructed similarly in principle to maze 1 but of hard fibre board, and employed the pattern of Foster's maze A. The same stylus was used in both mazes. As punishment a shock of 27 volts A.C. was used, and the intensity kept constant for all subjects receiving punishment in this experiment. The subject, with one hand on the metal plate at the side of the maze, completed the circuit when the stylus, which was held between the thumb and two fingers of the other hand, came in contact with the end of a blind alley. The shock was felt in the hand grasping the stylus. The maze, the entire table top, and the experimenter were screened from the subject's view as described in the earlier study.

Subjects.—One hundred and ten university students from classes in Introductory Psychology served as subjects. Of these, one hundred, 50 men and 50 women, completed the experiment. Ten records were discarded because of failure to meet certain criteria, as explained below. None of the subjects had had previous experience in maze learning.

Procedure.—The problem to be learned was mastered by the whole method and at a single sitting. The same individual acted as experimenter throughout the experiment and all records were obtained in the same room under as nearly the same conditions as possible. Some of the subjects, with both the original and the interpolated mazes, learned with punishment for errors, and, some learned under non-punishment or normal conditions. The criterion of mastery, in both learning and relearning, was that of three successive trials without error.

The directions were given orally and were slightly different for the two conditions, n that the members of the punishment group were informed that they would receive a slight shock when the stylus came in contact with the end of a blind alley. Aside from the differences resulting from punishment for errors and the instructions to this effect, every effort was made to keep the conditions constant for all experimental groups.

The subjects, when invited to take part in the experiment, were not informed of the nature or purpose of the experiment, nor were they told that they would have to come more than the one time. After the allotted interval had passed, the instructor told the subjects that there was another part of the experiment and asked them to

return for this part. Although this plan resulted in the loss of ten subjects, whose records had to be discarded because they failed to return, it had an advantage in that it eliminated possible rehearsing by the subjects during the intervening period.

Schedule of Groups.—The one hundred subjects were divided into five groups of twenty each. Each group consisted of ten men and ten women. The program followed by each group is indicated in Table 1.

TABLE 1

PROGRAM OF THE CONDITIONS OF LEARNING FOR EACH GROUP

Group	Conditions of Learning the First Maze	Interval	Conditions of Learning the Interpolated Maze	Interval	Relearning
NN...	*N*	3 weeks	*N*	3 weeks	*N*
NS....	*N*	3 weeks	*S*	3 weeks	*N*
SN....	*S*	3 weeks	*N*	3 weeks	*N*
N.....	*N*	6 weeks interval			*N*
S.....	*S*	6 weeks interval			*N*

The letters *S* and *N* refer to the conditions under which the problem was learned. The letter *S* is used to designate punishment (*shock*) and *N* to designate non-punishment (*normal*) learning. All groups relearned without punishment. The time interval between learning and interpolation and between interpolation and relearning, as well as the longer intervals between learning and relearning, were strictly adhered to with all subjects; when the interval would have had to be varied more than two days the cases were discarded.

RESULTS

1. *The Influence of Punishment During the Learning of the Stylus Maze upon (a) the Ability to Master the Maze, and (b) the Retention of the Maze Solution*

(*a*) As indicated in a preceding section, all subjects learned maze 1 as the first part of their work. Of the one hundred subjects used in the experiment, sixty learned maze 1 under normal conditions (groups *N*, *NN*, and *NS*), and forty mastered maze 1 under punishment conditions (groups *S*, and *SN*). A comparison of the records of the latter with those of the former will show the influence of administering electric shock as punishment for errors made while learning the maze. In Table 2 are given the averages of the learning scores, their differences, and the probable errors of the differences, for both the normal and the punishment conditions.

It is evident from these data that electric shock punishment for errors had a beneficial effect upon the learning scores. The punishment groups, in terms of the averages,

TABLE 2

LEARNING DATA FOR MAZE 1

Group	No. of Subjects	Trials Ave. No.	Trials P.E.	Time Ave. Sec.	Time P.E.	Errors Ave. No.	Errors P.E.
N	20	37.25	5.20	1279.0	116.78	416.45	46.34
S	20	24.25	3.71	1081.3	96.77	270.9	28.26
Difference		13	6.3	197.7	151.6	145.55	54.2
Percent decrease of *S* under *N*		*35.1*		*15.4*		*34.9*	
N, NN, NS	60	36.45	2.82	1271.16	74.21	395.03	32.31
S, SN	40	23.13	2.01	1122.5	71.35	275.15	17.62
Difference		13.32	3.46	148.66	102.9	119.88	36.8
Percent decrease of *S* under *N*		*36.5*		*11.6*		*30.3*	

required less total time, fewer trials and fewer errors than did the sixty subjects who learned under normal conditions. The punished individuals averaged 36.5 percent fewer trials, 30.3 percent fewer errors, and 11.6 percent less total time than did the unpunished individuals. In general, the differences obtained in this study confirm those obtained in the previous experiment with punishment in the stylus maze (1).

(*b*) The influence of the punishment during learning upon the retention of the maze activity over the interval of six weeks may be obtained by a comparison of the relearning, recall, and saving scores of the punished group (*S*) with the non-punished group (*N*). These data, and their probable errors, are presented in Table 3. Whether punishment during learning has a beneficial or detrimental effect upon retention seems to be in part a function of the method of measurement employed. In terms of the *recall* scores (scores on the first relearning trial) the influence of the punishment during learning was to *increase* the number of errors made by 46 percent and the time required by 38 percent. However, to *relearn* the problem to the same criterion as in original learning, the punishment individuals required fewer trials, made fewer errors and took less time than did the control group. The differences in the relearning scores indicate that the effect of the punishment was to *decrease* trials by 46 percent,

errors by 24 percent, and time by 41 percent. When the problem is mastered under normal conditions the amount retained over the interval of six weeks in terms of *percent saving* is 70.33 percent for trials, 93.97 percent for errors, and 82.99 percent for time. The corresponding retention values for the punishment condition are 75.46 percent for trials, 92.98 percent for errors, and 88.24 percent for time. The punished individuals show a greater percentage of saving for the interval in terms of trials and time, while in terms of errors the two groups are approximately equal.

From the standpoint of retention then, punishment during learning is beneficial when retention is measured in terms of relearning or in terms of percentage of saving over the original learning, though its effect is detrimental when retention is measured in terms of recall scores.

The lack of agreement between the recall scores and the other criteria is probably quite largely a function of the conditions employed in this experiment rather than a result of differences in the methods of measuring retention *per se*. In support of this view it will be remembered that the subjects in the punished group (*S*) relearned the maze under normal or non-punishment conditions, as did all the groups; whereas during the original learning they had received an electric shock each time they came into contact with the end of the blind alley. Hence, the conditions under which relearning took place were more nearly identical with those of the original learning for the normal group (*N*) than for the punishment group (*S*). Poorer records during the first relearning trial (recall scores), might conceivably be due to the absence of factors, evidently helpful cues, which were present during learning. Such a variation, absence of shock during relearning when it had been present during learning, might be a slightly disturbing factor to which, however, the subjects could soon adjust themselves, probably during the first relearning trial. It is probable then that the recall scores would be more detrimentally affected than the relearning scores by the introduction of such a variation between the relearning and the learning conditions. No group in the

present experiment relearned the maze under punishment conditions after having learned it with shock and so data in proof of this point are not available. The fact remains, however, that recall under normal conditions six weeks after the mastery of the maze under punishment conditions is poorer than recall under normal conditions when the original learning occurred without punishment; while relearning under normal conditions is better when the original learning occurred under punishment conditions than when no such punishment was introduced during the original mastery of the problem.

2. *The Influence of a Second Maze Problem upon the Retention of the First When the Latter Is Learned Under (a) Normal Conditions, and (b) Under Punishment Conditions*

(*a*) The influence of the interpolated maze upon the retention of the original problem is obtained by the comparison of the recall, relearning and saving scores of group *N* with those of group *NN*. Group *N* relearned the original problem six weeks after mastery without learning a second maze during the interval. Group *NN* learned a second maze three weeks after learning the original problem and three weeks before relearning the latter. The retention scores for the two groups appear in Table 3; the differences between the two groups and the probable errors of the differences appear in Table 4. These data show that the acquisition of the second maze during the interval interfered with the retention of the first problem by all the criteria except relearning trials and percent saving for trials. The effect of the learning of the interpolated maze upon the retention of the previously learned maze solution was: an increase of 61.19 percent in the amount of time spent in relearning and of 336.05 percent in the number of errors made in relearning; in terms of saving scores, a 16 percent less saving in errors and 6 percent less saving in time; in terms of recall scores, an increase of 1238 percent in errors and of 385 percent in the time required.

Thus, a comparison of the retention scores of the two groups indicates that the interpolated problem had a slightly

TABLE 3
RETENTION DATA FOR MAZE I

Group	No. of Sub.	Relearning						Percent Saving			Recall			
		Trials		Time		Errors					Time		Errors	
		Ave. No.	P.E.	Ave. Sec.	P.E.	Ave. No.	P.E.	Trials	Time	Errors	Ave. Sec.	P.E.	Ave. No.	P.E.
N....	20	11.05	1.19	217.5	30.32	25.1	3.23	70.33	82.99	93.77	30.06	2.07	4.85	.95
S.....	20	5.95	1.17	127.1	19.19	19	3.12	75.46	88.24	92.98	41.5	4.89	7.1	1.4
NN...	20	8.9	1.18	350.6	62.14	109.45	27.22	77.69	76.61	77.69	146	43.78	64.9	21.94
SN...	20	8.4	.99	209	23.46	48.2	8.85	61.81	82.04	82.74	87.05	15.31	31.95	7.3
NS...	20	9.95	1.75	219.7	24.88	53.25	9.58	69.09	78.78	80.84	74.3	11.94	26.9	6.05

facilitating effect upon the retention of the first problem when retention was measured by the number of trials required to relearn and by the saving score for trials, but a marked interference effect upon the retention of the first problem in terms of errors and time of relearning, errors and time of recall, and in the savings scores for errors and time. More than four times as many errors occurred in relearning after interpolation and more than thirteen times as many in recall; about one and two thirds times as much time was necessary in relearning and almost five times as much time was necessary in recall. Statistical measures of reliability indicate that the errors of relearning and errors and time of recall represent the most significant differences between the groups; although in no instance is the difference as much as four times the probable error of the difference, the ratios for these measures are sufficiently high that the chances are greater than 90 in 100 that the direction of the differences obtained would be confirmed by subsequent work.

(*b*) The second problem to be considered under the above heading is the effect of punishment during the original learning upon the amount of retroactive inhibition. What is the influence of punishment for errors during the original learning upon the resistance of this maze solution to the interference resulting from the learning of a second maze problem? Two comparisons are involved here; first, the differences in the retention scores between groups *N* and *NN* must be compared with the differences in the retention scores between groups *N* and *SN*; second, the differences in the retention scores between groups *S* and *SN* must be compared with the differences between groups *N* and *NN*.

The influence of the interpolated maze learning (normal condition) upon the retention of the first learned maze, when the latter was mastered under normal conditions, as just reviewed in part (*a*) above, was a beneficial effect of 19.4 percent in the number of trials required to relearn, a detrimental effect of 336 percent in terms of relearning errors, and a detrimental effect of 61 percent in the time required to relearn, over the corresponding values obtained when no

TABLE 4
GROUP DIFFERENCES IN RETENTION

Groups Compared	Relearning						Percent Saving			Recall			
	Trials		Time		Errors					Time		Errors	
	Ave. No.	P.E. Diff.	Ave. Sec.	P.E. Diff.	Ave. No.	P.E. Diff.	Trials	Time	Errors	Ave. Sec.	P.E. Diff.	Ave. No.	P.E. Diff.
NN–N	*2.15	1.68	133.10	69.09	84.35	27.39	*7.36	6.38	16.28	115.94	48.44	60.05	21.96
Percent increase of *NN* over *N*	*19.4		61.19		336.05					385.69		1,238.14	
SN–N	*2.65	1.54	*8.15	38.33	23.10	9.10	8.52	.95	11.23	56.99	15.44	27.10	7.36
Percent increase of *SN* over *N*	*23.9		*3.91		92.03					185.59		558.76	
SN–S	2.45	1.53	81.90	30.30	29.20	9.38	13.65	6.20	10.24	45.55	16.07	24.85	7.43
Percent increase of *SN* over *S*	41.18		64.44		153.68					109.76		350.14	
NS–N	*1.10	2.11	2.20	39.22	28.15	10.11	1.24	4.21	3.13	44.24	12.11	22.05	6.12
Percent increase of *NS* over *N*	*9.95		1.01		112.15					147.17		454.64	

* This difference favors the control over the interpolated group, indicating facilitation rather than interference.

second problem is interpolated. The influence of the same interpolated maze upon the retention of the original maze (see Table 4), when the latter was mastered under punishment conditions, was, in terms of relearning scores, a decrease of 23.9 percent in the number of trials, a decrease of 3.9 percent in the amount of time required, and an increase of 92 percent in the number of errors made. In other words, when the first problem was learned under normal conditions the interpolated problem had, in terms of relearning scores, a slightly beneficial effect upon trials and a decidedly detrimental effect upon both errors and time; but, when the first problem was learned under punishment conditions, the effect of the interpolated maze was slightly beneficial in both trials and time and much less detrimental as regards errors than when no punishment was administered during the learning.

In terms of recall scores the interpolated problem had a greater interference effect upon the retention of the first problem when the latter was learned under normal conditions than when punishment was given. There was marked retroactive inhibition in both cases, but the relative amount was reduced approximately one half by administering punishment during the original learning.

In terms of percent saving, a comparison of the differences between groups *SN* and *N* with those between *NN* and *N* shows that greater retention values, hence less retroactive inhibition, occurred when punishment was not administered during learning. It should be remembered in this connection that saving scores are markedly influenced by differences in the magnitude of the original learning scores. Since punishment for errors during learning, as indicated in Table 2 above, decreases the amount of time and the number of errors and trials required to meet the criterion of mastery, saving scores might be expected *a priori* to be greater for the *SN* group.

Although these results for groups *N* and *SN* do seem to indicate that a maze solution learned with punishment is less susceptible to interference than one learned under normal conditions, the data thus far presented are not decisive. The fact has been brought out previously that punishment

during learning improves retention as measured by relearning. Since this is true, it is quite impossible to say how much of the improvement of the *SN* group over the *NN* group in comparison with the *N* group is correctly ascribed to improved retention and how much to increased resistance to interference —if it is admitted that there may be a difference between these two conditions. Furthermore, as has just been pointed out, direct comparison of the saving scores for the *SN* group with those for the *N* group results in spuriously increased differences.

To avoid these objections a further comparison has been made of the differences in retention scores between groups *S* and *SN* with those between groups *N* and *NN*. The differences between the retention scores of the subjects in group *S* and those in group *SN* indicate the detrimental effect of the interpolated problem upon the retention of the original maze when the latter was learned under punishment conditions. A comparison of these differences with those between *N* and *NN* (see Table 4) shows the influence of shock during learning upon the susceptibility of the problem learned to the detrimental effect of the second problem. When the original maze was learned under normal conditions, the effect of the interpolated maze upon retention, as already shown, was slightly beneficial in terms of relearning trials and saving scores for trials, but decidedly detrimental in terms of all other measures. When the original learning occurred under punishment conditions, the interpolated problem had a detrimental effect upon retention according to all the measures used, but this detrimental effect was much less than that which occurred when the original learning was under normal conditions. This statement is true for all measures except relearning trials and saving score for trials, the two measures which showed slightly beneficial effect of interpolation under the *NN–N* condition. The decrease in the interference when punishment was given during the original problem over that which occurred when no punishment was given is especially marked in errors of relearning, and in errors and time of recall.

In other words, according to all measures the retention of the maze solution learned under punishment conditions showed interference resulting from the acquisition of a second maze solution during the interval between learning and recall; but, with the punishment group the interference which resulted from the interpolated problem was less according to all measures except two, than that which occurred when learning took place under normal conditions. Under both normal and punishment conditions the interpolated problem had its greatest detrimental effect in terms of relearning errors and of errors and time of recall. In terms of these measures the detrimental effect of the second maze was less than one half as great when original learning occurred under punishment rather than under normal conditions.

3. *The Influence of a Second Maze Problem Upon the Retention of the First When Punishment Is Administered During the Interpolated Learning*

To determine the influence of administering punishment for errors during the mastery of the interpolated maze upon the detrimental effect which it exerts upon the retention of the first problem it is necessary to compare (*a*) the differences between the retention scores of group *NN* and group *N* with the differences between group *NS* and group *N*; and (*b*) the raw retention scores of the *NN* group with those of the *NS* group. A direct comparison here is justified since the difference in the experimental conditions between *NN* and *NS* is that of punishment during the interpolated learning so that group *N* serves as a control for both.

(*a*) An inspection of the data in Table 4 showing *NN–N* and *NS–N* indicates that the detrimental effect of the interpolated maze upon the retention of the first maze is much greater when the interpolation is mastered under normal conditions than when learned with punishment. When the interpolated maze is learned under normal conditions (group *NN*) its detrimental influence may be determined by a comparison of the resulting retention scores with those of group *N*. Such a comparison shows an increase of 336 percent in

relearning errors, of 61 percent in relearning time, of 1238 percent in recall errors, and of 385 percent in recall time; as well as a decrease of 16 percent in the saving score for time. When the interpolated maze is learned under punishment conditions (group *NS*) a similar comparison with group *N* gives the following results: an increase of 112 percent in relearning errors, of 1 percent in relearning time, of 454 percent in recall errors, and of 147 percent in recall time; while the decrease in saving scores amounted to 1.2 percent for trials, 3.1 percent for errors, and 4.2 percent for time. In terms of these measures of retention, then, the differences between *NN* and *N* are considerably greater than those between *NS* and *N*. This comparison indicates that the administration of punishment during the learning of the interpolated problem tends to decrease the retroactive inhibition produced. The differences in the saving scores for trials, as found in most of the comparisons above, show greater detrimental effect in the punishment group.

(*b*) A comparison of the relearning records of the *NN* group directly with those of the *NS* group, Table 3, shows that the latter relearned the first problem with 51 percent fewer errors and in 37 percent less time, though they required a slightly greater number of trials. In terms of recall scores the *NS* group made 59 percent fewer errors and required 49 percent less time. The *NS* group, in terms of percent saving, shows a slightly greater retention in errors and in time, but less in trials, than did the *NN* group. The most significant differences occur in errors and time of relearning and in errors and time of recall. These differences indicate that retention of a previously learned maze solution, as measured by raw scores in relearning and recall, is more efficient when the problem interpolated between the original learning and the retention test is mastered under punishment rather than normal conditions.

4. *The Relative Influence of Punishment, Upon the Amount of Retroactive Inhibition, When Administered During the Learning of the First Maze as Compared with Its Effect When Introduced During the Learning of the Interpolated Maze*

In sections 2 and 3 above it was found that the disintegrative effect of the interpolated maze upon the maze previously learned is decreased when punishment for errors is given during the learning of the first maze, and that it is also decreased when punishment is introduced in the learning of the interpolated maze. One other problem remains: is the decrease in the amount of retroactive inhibition greater when the maze to be relearned has been learned with punishment, or, when the interpolated maze has been learned with punishment? This relationship may be determined by comparing the differences between *NN–N* and *SN–S* with those between *NN–N* and *NS–N*. The data are shown in Table 4. These differences between the two punishment conditions are much smaller than those considered in the preceding comparisons and furnish indications only. In terms of relearning and saving scores punishment during the learning of the interpolated maze had the greater effect; whereas, in terms of recall scores, the decrease in the amount of retroactive inhibition was greater when punishment was administered during the learning of the first maze.

A comparison of the raw relearning and recall records of the *NS* group with those of the *SN* group in Table 3 shows that the two groups have just the reverse of the relation indicated above. Relearning records were better when the first maze was learned under punishment; while the recall scores were better when the punishment was introduced during the learning of the interpolated maze.

If we regard as a single problem the learning of the first maze and the relearning of this maze six weeks later, that is, three weeks after having learned a second maze, then we may compare the groups in terms of the sums of their raw learning and relearning scores. Such a comparison will indicate the most efficient condition (punishment on first problem, punish-

ment on interpolated problem, or no punishment at all) under which a group is able to learn a maze to a definite criterion of mastery and then to relearn it six weeks later to the same criterion. The most efficient group would be that one which learned and relearned the problem in the shortest time with the fewest trials and the smallest number of errors. In general, the groups rank as follows, in decreasing order of efficiency: *SN*, *NS*, and *NN*. Group *SN* was, of course, the only group which received punishment during the original learning, and probably ranks first because of this fact. The results, already discussed, indicate that punishment for errors during learning reduces markedly the amount of time and the number of trials and errors required to learn (Table 2). In relearning the maze group *NS* made better records than group *NN* except in trials but poorer records than group *SN* (Table 3). These results further substantiate the previous finding that punishment for errors during maze learning reduces both the initial scores in learning and the amount of retroactive inhibition as shown in the retention criteria.

In addition to the foregoing analyses the obtained data make possible certain further studies, which, though not directly relevant to the main thesis of this investigation, are yet presented here as interesting by-products of this research. In the following three sections we shall discuss, first, transfer as shown in the mastery of the interpolated maze, second, sex differences in learning, and third, the relation of learning scores and intelligence test scores.

5. *The Influence of Punishment for Errors During the Learning of the First Maze Upon the Mastery of the Interpolated Maze*

In Table 5 are given the interpolated maze learning records. A comparison of these records indicates: first, that punishment for errors in the first maze facilitated the learning of the second maze slightly more than did the learning of the first maze without punishment; and second, that, following normal learning in the first maze, punishment facilitated the mastery of the second maze in two of the criteria though not

in terms of time. In the absence of a group which learned maze 2 without having learned a previous maze, these differences must be considered as suggestive only. Accordingly, we do not feel justified in making any conclusive statements with regard to the influence of punishment upon transfer.

TABLE 5

INTERPOLATED MAZE LEARNING RECORDS

Group	Trials		Time		Errors	
	Ave. No.	P.E.	Ave. Sec.	P.E.	Ave. No.	P.E.
NN.....	27.25	2.69	1649.5	135.78	407.3	34.06
SN.....	27.45	4.04	1540.4	122.46	384.9	38.24
NS.....	22	2.75	1680.9	161.50	339.6	29.86

6. *Sex Differences in Learning, With and Without Punishment*

Since a large number of subjects were used in this experiment and since half of the subjects were men and half women in every group, a brief consideration of the sex differences is given here. In Table 6 are given the mean learning scores of men and women with and without punishment, together with probable errors, coefficients of variability $\left(\frac{100\sigma}{\text{Average}}\right)$, and the ratio of the scores of the female to the male (F/M). Although none of these differences is statistically reliable, a number of them are significantly large and sufficiently consistent to warrant mention and brief consideration.

TABLE 6

SEX DIFFERENCES IN LEARNING MAZE 1

Sex	Cond. of Learn.	No. of S.	Trials			Time			Errors		
			Av. No.	P.E.	V.	Av. Sec.	P.E.	V.	Av. No.	P.E.	V.
Male....	*N*	30	28.3	3.5	100.5	941.17	81.57	70.1	282.6	29.6	85.0
Female.	*N*	30	44.6	4.2	77.1	1634.50	125.39	62.3	507.46	54.0	86.6
Ratio...	(F/M)		1.57			1.79			1.74		
Male....	*S*	20	15.65	1.1	46.7	790.3	51.12	42.9	229.75	23.8	68.7
Female.	*S*	20	30.60	3.5	76.0	1454.7	100.07	45.6	320.55	24.1	49.8
Ratio...	(F/M)		1.96			1.39			1.84		

Both with and without punishment the men made much better learning records by all criteria. Punishment increased the advantage of the men in all criteria except that of errors. For both men and women by all criteria punishment facilitated learning and decreased variability. Without punishment the men are more variable in all criteria except that of errors, where the coefficient of variability is very slightly higher for the women. With punishment the women are more variable in all criteria except that of errors, and here the men have a considerably higher coefficient.

7. *Correlations of Maze Learning with Test Intelligence*

In Table 7 are given the correlations of maze learning scores with scores on the Otis Group Intelligence Scale, Form A. The correlations are very much higher with punishment learning.

TABLE 7

CORRELATIONS WITH OTIS SCORES

Conditions of Learning	Number of Cases	Trials		Time		Errors	
		r	P.E.	*r*	P.E.	*r*	P.E.
N.....	56	−.149	.088	.308	.081	.214	.085
S.....	33	.419	.096	.685	.062	.796	.043

DISCUSSION OF RESULTS

The learning of a second maze has a detrimental effect upon the retention of a previously learned maze solution by all criteria except that of trials. There is evidence from all of the groups in support of this conclusion. Trials of relearning show no deleterious results of interpolation in the non-punishment groups, but the reliability as indicated by the probable error is less for this criterion than for any of the others which were used.

Retroactive inhibition such as that which occurred in this experiment can hardly be explained by the perseveration hypothesis of Müller and Pilzecker. Even its most ardent defenders could scarcely maintain that the after-effects of neural activity persisted for three weeks, which was the

interval between the original learning and the interpolation in this experiment. The results can be more readily explained in terms of a transfer theory. It is probable that the learning of the interpolated maze resulted in some disintegration of the original maze pattern and a recombination of its elements into another pattern, or it may have resulted in the establishment of different responses to very similar stimuli. Hence, recall is poorer. It is also probable that the interference may be present at the time of recall when the individual is confronted with the original stimulating situation. In this condition it is a question whether the activities acquired in learning the first or those acquired in learning the second will be aroused when the original stimuli are presented. Whichever activity is aroused is probably dependent upon the strength of the associative connections (or conditionings), and upon the exactness of the discriminations between the stimuli.

Apparently punishment acts as motivation to facilitate learning, interpolated learning and relearning. An electric shock of 27 volts A.C. seems to be far more potent in establishing an avoiding response than the mere blocking at the end of an alley.

Punishment facilitates learning and in addition decreases variability. It would seem from the correlations with the intelligence test scores that the effort expended to reach a maze solution varies more than the actual ability to learn. When the motivation is increased, the subjects presumably make a greater effort and learning more nearly approaches the best of which the subject is capable.

The learning of a second maze has a more pronounced detrimental effect upon the retention of a previously learned maze solution when neither maze has been learned with punishment. There are several possible reasons why punishment during either the first learning or during the interpolated learning decreases the amount of retroactive inhibition. The factors to be mentioned are not unrelated and all may have operated to some extent.

1. Increased motivation and emotional accompaniment of

punishment. The fact that an electric shock was administered as punishment for certain errors was not regarded as an insignificant factor in this experiment. Most of the subjects seemed to regard it, as indicated by their verbal behavior, as a rather unpleasant accompaniment of making an error. Apparently punishment facilitated learning by arousing a more active attitude on the part of the subject and by increasing the motivation to learn. It is possible that a maze problem learned with a greater amount of coöperation and reënforcement from the individual's other motor mechanisms has greater retention value, and consequently less susceptibility to the disintegrating effect of a second similar problem. It may have greater retention value because the presence of a more active attitude and emotional reënforcement may result in a more definite integration of the acquired activity. Also, if a maze solution is more highly integrated, as a result of being acquired under punishment conditions, it may exert a less detrimental effect upon a previously learned maze solution than when the interpolated maze solution has a lesser degree of integration as a result of learning under normal conditions.

2. Decreased similarity. Previous studies indicate that the amount of retroactive inhibition is largely a function of the degree of similarity between the original and the interpolated learning. Administering punishment for errors, either during the original learning or during the interpolated learning, decreases the degree of similarity between the two problem situations as compared with normal learning conditions present during both the original and the interpolated learning. It is possible that decreasing the similarity between the two learning situations, by the introduction of such a variable as punishment in one but not in the other, also decreases the detrimental influence exerted by the interpolated problem. While it is unlikely that all of the differences in our results are due to this factor alone, no group in this experiment learned the first problem, the second, and relearned the first under punishment conditions and so we do not have the data for the necessary comparison.

3. Decreased maze activity. As indicated in the presentation of the results, punishment for errors results in quicker learning of the first maze and quicker learning of the interpolated problem. Hence the punished individuals, on the average, spent less time in the maze and made fewer errors than did the nonpunished individuals. It is possible that the amount of maze activity indulged in during the retention interval may be causally related to the interference which it exerts upon a previously acquired maze solution in such a way that a significant decrease in the amount of such interpolated activity is followed by better retention than would occur without such a decrease. It is also possible that decreasing the mistakes and the unnecessary activity in the original learning may result in a greater difference between the functional strength of the acts 'selected' and that of the acts 'eliminated' than would be present were no punishment used, and, consequently, better retention during the intervening interval with greater resistance to interference from a second maze solution. By this influence upon the functional strength of the associations involved decreased maze activity might operate as another factor determining the degree of the integration of the whole solution.

Conclusions

1. Administering punishment for errors during the mastery of a stylus maze resulted in decidedly quicker learning by all criteria, *i.e.*, a marked reduction in the amount of time, the number of errors and trials required to meet the criterion of learning.

2. Punishment for errors during learning resulted in greater retention of the maze solution for an interval of six weeks, when retention is measured by the relearning method; though, when measured by recall scores, retention was more efficient for a maze learned without punishment.

3. The mastery of a second maze three weeks after learning the first and three weeks before the retention test of the latter resulted in marked retroactive inhibition by all criteria except trials.

4. The detrimental effect of the second maze upon the retention of the first was less than one half as great, except in terms of trials, when original learning occurred under punishment as when it occurred under normal non-punishment conditions.

5. The amount of retroactive inhibition resulting from the acquisition of a second maze is also reduced approximately one half, except in terms of trials, when punishment is administered for errors during the learning of the interpolated maze problem.

References

1. Bunch, Marion E., The effect of electric shock as punishment for errors in human maze-learning, *J. Comp. Psychol.*, 1928, **8**, 343–359.
2. Carr, H. A., The influence of visual guidance in maze learning, J. Exp. Psychol., 1921, **4**, 399–417.
3. Harden, Luberta M., The effect of emotional reactions upon retention, *J. Gen. Psychol.*, 1930, **3**, 197–221.
4. McGeoch, John A., On the term retroactive inhibition, *Amer. J. Psychol.*, 1930, **42**, 455–457.
5. Skaggs, E. B., Further studies in retroactive inhibition, *Psychol. Monog.*, 1925, **34** (No. 161), Pp. 59.
6. Tolman, E. C., Retroactive inhibition as affected by conditions of learning, *Psychol. Monog.*, 1917–18, **25** (No. 107), Pp. 50.
7. Webb, L. W., Transfer of training and retroaction, a comparative study, *Psychol. Monog.*, 1917, **24** (No. 104), Pp. 90.
8. Whitely, Paul L., The dependence of learning and recall upon prior intellectual activities, J. Exp. Psychol., 1927, **10**, 489–508.

TRANSFER OF RESPONSE IN HUMAN SUBJECTS[1]

BY HAROLD GULLIKSEN

From the Psychological Laboratory of the University of Chicago

A wide variety of investigations have appeared in the literature under the general head of transfer. They differ both in the factors varied experimentally, and in the criterion of transfer. As to the factors varied experimentally, sometimes the response has been kept constant, and the stimulus varied; sometimes the reverse has been done. Sometimes the stimulus and response have both been kept constant and the context altered. In some experiments both the stimulus and response have been altered, as has been the case in many of the experiments on transfer in education, and on transfer in maze learning. For instance, when a rat trained on one maze is tested on other mazes, the mazes (stimuli) are different, and the adjustments made (responses) are different in each case. This variation of several factors at once makes it difficult to analyze and interpret the data.

As to the criteria of transfer, two different ones have been used. One criterion is change in the subject's ability to learn new things. This is used for example in most of the animal experiments demonstrating transfer on the basis of relative rather than absolute factors in the stimulating situation. In most of these experiments punishment and reward are continued in the testing series just as they are in the training series, so that the animal is *learning* how to respond in the test series as well as in the training series. The other criterion is change in the subject's stimulus-response connections. This criterion is used for example in a few of the animal experiments on transfer. For example in the work of Klüver, of Lashley, and of Pavlov, the differential punishment and reward are not given during the test series, but only during

[1] This experiment was carried on under the direction of Dr. Harvey A. Carr and I wish to express my appreciation for his valuable criticism and advice.

the training series, so that during the test series the animal does not have the same opportunity to learn how to respond, that he has during the training series. With the first technique it is impossible to tell to what extent the results are due to increased rapidity of learning of new problems, and to what extent to an immediate transfer of a given response from the old stimulus to the new one. By using the second criterion, the factor of immediate transfer of a given response from the old stimulus to the new one, is isolated.

In this experiment we are dealing only with the type of condition in which the context and response are kept constant and the stimulus is altered; and the criterion of transfer is the change brought about in the functioning stimulus-response mechanisms in the organism. It has been shown that when an organism has been trained to give a certain response to one stimulus, it sometimes without further training gives the same response to certain other stimuli. For example Watson found that a child trained to show fear whenever a rat was presented, would, without any further training, show fear to a rabbit, a fur muff, a pigeon, and a Santa Claus mask. None of these objects had formerly called forth the fear response. Furthermore this transfer of response did not extend to *all* other stimuli. It was not given to the blocks the child played with, to the room, to the table, nor to the experimenter. We see here an illustration of the fact that when an organism has been trained to give a response to one stimulus it may without further training give the same response to certain other stimuli, but not to all other stimuli. Dr. Watson, however contented himself with simply stating the specific facts shown by his experiment without attempting to find the general *differentiating* principles involved. Furthermore he treats his data as if the test stimuli used fell into two distinct classes; 1. stimuli to which the response did transfer, and 2. stimuli to which the response did not transfer. Yum, on the other hand, working on 24 hour recall of paired associates, found evidence not for an all-or-none transfer, but a relationship between similarity of the test stimuli to the training stimuli, and the probability of recall. Klüver working with

monkeys, and Pavlov with dogs agree with Watson, in that they report their findings in terms of transfer, or no transfer, instead of in terms of the degree of transfer. Watson then reports transfer on an all-or-none basis for infants, and this agrees with the reports of Pavlov for dogs, and of Klüver for monkeys; while Yum working with adult human subjects reports that the degree of transfer depends on the degree of similarity and is not an all-or-none matter. It is important to know whether transfer is an all-or-none affair, or is exhibited in varying degrees. This is one of the problems that will be taken up in this paper.

The present investigation is an attempt to set up an experiment which will get at some of the differentiating principles in accordance with which the new stimuli will or will not call out the old response. A comparatively simple set-up was arranged for using adult human subjects under conditions more closely paralleling those generally used for animals. The subject was first trained to give two antagonistic responses, one to each of two classes of stimuli. Then with the context kept constant, other stimuli were presented which bore varying relations to the two stimuli of the training series. The subject was required to give either one of the two antagonistic responses to each of the new stimuli. The responses of the subject, and the relations between the training and test stimuli were then examined to determine what generalizations would enable one to predict which of the two antagonistic responses would be given to any test stimulus.

Training Procedure

The subjects were trained to react differentially to two different series of visual stimuli, namely two different types of geometrical angles. A representative sampling of the training stimuli used is shown in Plate I. There were two different series of stimuli used.

One group of subjects was trained, using series *X* (Plate I), to give one response to stimuli of type *x*1, and an antagonistic response to stimuli of type *x*2. There were 100 stimuli each of type *x*1, and *x*2. As can be seen from Plate I the different stimuli of series *X* are angles which vary in the length and thickness of the sides of the angle, but are all alike in that the point of each angle is directed toward the upper right hand corner. The difference between the members of type *x*1, and type *x*2 is that the stimuli of type *x*1 are all 100° angles, while the stimuli of type *x*2 are all 25° angles. So in series *X* the differential reaction of the subject is made on the basis of the *size* of the angle.

Another group of subjects was trained using series *Y* (Plate I), to give one response to stimuli of type *y*1, and an antagonistic response to stimuli of type *y*2. There were 100 stimuli each of type *y*1, and type *y*2. The members of series *Y* like those of series *X* are angles which vary in the length and thickness of the sides of the angle. The stimuli of series *Y* are all alike in that they are all 45° angles. The difference between members of type *y*1 and type *y*2 is that stimuli of type *y*1 are angles with the vertex directed toward the left, while type *y*2 is composed of angles with the vertex directed downward, so that in series *Y* the subject's differential reaction is made on the basis of the *direction* of the angle.

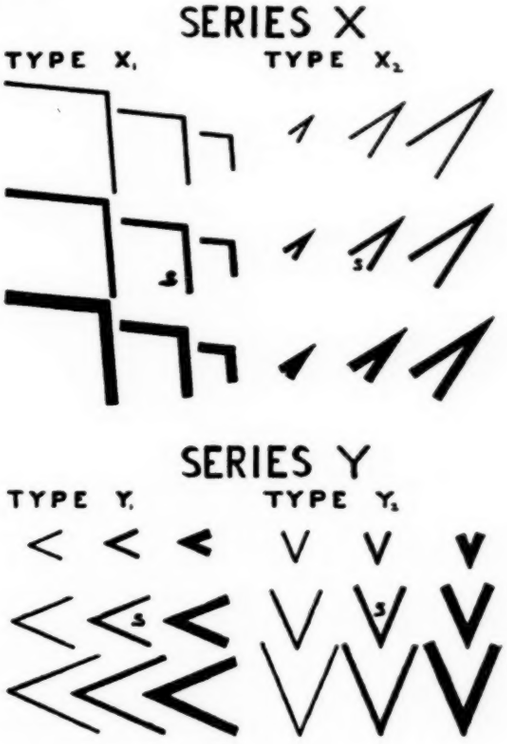


PLATE I

Each stimulus figure was drawn in india ink on a 5 x 6 inch card, $\frac{1}{16}$ inch thick, made of a special fiber board that is very tough and yet takes drawing or printing very legibly. These cards were then presented to the subject one at a time. The exposure period was about 1½ seconds, and the interval between exposures was about 1½ seconds. The cards were presented by a special type of tachistoscope designed to present long series of visual stimuli at any desired constant rate. This apparatus has been described in an apparatus note appearing in the *Journal of General Psychology*.

At each sitting which lasted about 15 minutes the subject was presented with and reacted to 200 of the training stimuli. Each subject was given four sittings, one a day at the same time on each of four successive days. The same group of 200 stimuli was presented to the subject at each sitting.

The differential response required of the subject was the movement of a small lever to the right for one type of stimulus, and to the left for the other type of stimulus.

The method of training is best explained by first giving the instructions that were presented to the subject.

This buzzer (indicating a buzzer hung just above the reaction lever) will sound at times during the experiment. You are to react to the stimuli appearing and dis-

appearing here (indicating the opening in which the cards were to be exposed to the subject) by moving this lever to one side or to the other, or by leaving it in the middle (illustrating by moving the lever). By reacting properly you can keep the buzzer completely silent. The buzzer should be silent both during the presentation of stimuli and between presentations of stimuli. Please do not talk about this experiment with anyone.

The apparatus was then set so that the buzzer would ring under the following conditions:

1. If the lever were moved to the right or to the left during the 1½ second interval between presentation of cards.

2. If the lever were left in the neutral position for over ¼ second after the presentation of any card.

3. If the lever were moved in the "wrong" direction (specified in more detail below) when a card was presented.

4. If the lever were held to the right or to the left for over ¼ second after the disappearance of any card.

This meant that in order to keep the buzzer silent, *i.e.* to react correctly, the subject had to learn to react in the following manner:

1. To keep the lever in the neutral position between presentations of cards.

2. To move the lever in the 'correct' direction within ¼ second after the presentation of each card.

3. To allow the lever to return to neutral within ¼ second after the disappearance of any card, and leave it there until the next card appeared. During the ¼ second allowed for reaction time, the buzzer would remain silent no matter whether the reaction lever were left in the middle, or pushed to the right or to the left.

It was possible that a constant error might be introduced into the results due to the fact that a movement to the right was easier for the subjects to make than a movement to the left, or vice versa. To control for this factor the following method was used. With some subjects the right side was 'correct' for stimuli of type *x*1, and the left side was 'correct' for stimuli of type *x*2; while for other subjects the right side was 'correct' for stimuli of type *x*2, and the left side was 'correct' for stimuli of type *x*1. The same reversal of direction of movement for two different groups of subjects was used as a control for those trained to discriminate stimuli of type *y*1 from stimuli of type *y*2.

This particular method of training was adopted because it could, with slight changes, be used on young children or on animals. The verbal instructions used could be eliminated by the substitution of punishment of some sort for the warning signal of the buzzer; the response was a manual one instead of a verbal one; a choice reaction was required; and the learning was simply through the trial and error method. All of these features though somewhat unusual[2] in the human work on Transfer, were adopted in this experiment because the method could then be used on young children or animals with only minor changes, and so genetic and comparative investigations could be made in a more direct manner than otherwise.

Test Procedure

The test stimuli used were also geometrical angles. They are all shown in Plate II, series A, and series B. These test stimuli always bore one of three possible general relationships to one or the other of the two types of stimuli used in each training series.

---

[2] Ruediger, Thorndike, Volkmann, Woodrow, Yum.

(1) Some were like one or the other of the two types of training stimuli in *size*; but exhibited different degrees of variation in *direction*. (2) Some were like one or the other of the two types of training stimuli in *direction*; but exhibited different degrees of variation in *size*. (3) Other test stimuli differed from the two types of training stimuli in both *size and direction*. The exact number and kind of variations used can be seen by an inspection of Plate II, and will be described in greater detail below, when results are given.

Method of presenting the test stimuli: In the usual work with human subjects the procedure has been to present the subjects with the training series until they have reached a certain criterion of learning and then inform them that they are now to be presented with a test series in which they will not be told whether they are right or wrong. Sometimes the instructions have included the information that some of the

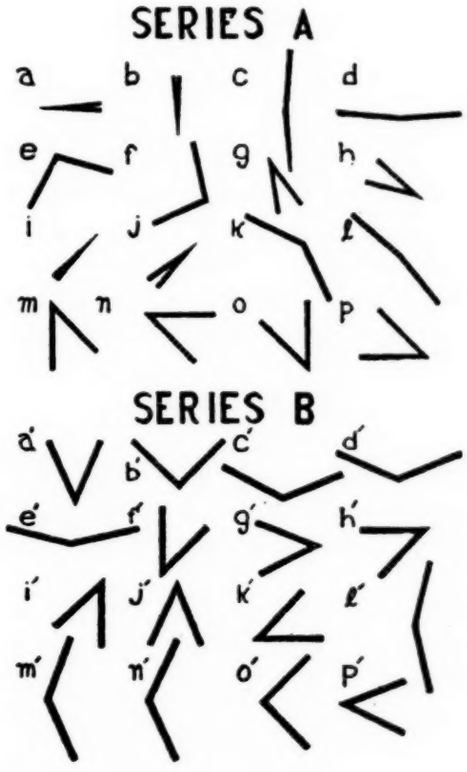


PLATE II

stimuli presented might not have been presented in the training series, but the subject is to react as seems best to him. Then the subject has been presented with the test series and his reactions recorded. With animals the same general procedure has usually been followed. The animal has been trained with punishment until it reached a certain criterion of learning and then has been presented with a long series of test stimuli without punishment.[3] In many of the experiments on animals punishment has been used in the test series.[4] In some work on transfer in rats I have found that during a test series of 30 successive trials the later reactions of the animal were frequently radically different from the first. In some cases it was noticed that even the

[3] Gayton, Pavlov, Washburn.

[4] Helson, Perkins, Wylie.

first one or two test trials consistently gave results different from those of the next eight test trials. This suggested the possibility that even though the animals had no instruction to the effect that they were being given a test series which differed from the training series in that punishment had been removed, nevertheless the difference between the training and test series was "noticed" sufficiently to cause the animals' reactions during the latter part of the test series to be different from their reactions during the earlier part. It seems that the first test trial made when the subject is all set to react to one of the regular training series is more indicative of the real transfer effect of that training series than the later test trials where one has not only the influence of the training series but also the influence of the preceding part of the test series. Therefore the customary method of presenting the test stimuli in a group after a certain criterion of learning had been reached was abandoned and the following procedure adopted.

The items of the test series were interspersed among the items of the training series at certain determined points, so that as far as the subject was concerned there was no difference between the stimuli of the test series and the stimuli of the training series. After a little preliminary experimentation it was found that most of the subjects had the training problem fairly well mastered after the presentation of about 40 of the training stimuli, so the test and training stimuli were arbitrarily arranged in the following order. Forty cards of the training series were presented, then one card of the test series, then 10 cards of the training series, then one card of the test series, then 10 more of the training series, and one of the test series, and so on, the series finally ending with the presentation of 10 cards of the training series. Two hundred training stimuli and 16 test stimuli were presented to the subject in this manner at a single sitting. With this method any two test stimuli were separated by 10 presentations of the regular training stimuli.

While this method has the disadvantage of introducing the various test stimuli at different stages of the learning process, it was thought worth trying because it minimizes the possibility of the results being altered by the discovery of the difference between the training and test series.

When the test stimuli were presented the cards themselves automatically adjusted the apparatus so that the buzzer would sound (¼ second after the presentation of the card) only if the reaction lever were left in the neutral position. Movement of the lever either to the right or to the left would keep the buzzer silent. The object was to see what response the subject would give to the test stimuli, and to discover how that response would be influenced by the type of training given.

In order to discover the difference between the test and training series when this procedure was used the subjects would have to notice that about every 10th card was different from the rest and that for those cards it did not matter in which direction the lever was moved. Furthermore they had only 1½ seconds to discover that the lever could be moved in either direction without the buzzer sounding. At the close of the experiment the subjects were asked to describe their solution of the problem. The question was framed as follows: "If you were instructing some one else how to react to this situation so that he would not have to go through the same trial and error process you did, what would you tell him?" Whenever a subject showed in his reports that he had discovered that the direction of movement was indifferent for some of the stimuli, his results were discarded. The results of nine subjects were discarded on this basis.

The subjects used were 72 students of elementary and advanced psychology in the University of Chicago. Three experiments were made, each with 24 subjects.

In the first experiment the subjects were trained on series *X* (Plate I) and tested with series *A* (Plate II). The 24 subjects used in this group were divided into two groups of 12 each, one group being trained to move the lever to the right for stimuli of type *x*1 and to the left for stimuli of type *x*2; the other group being given just the opposite training.

In the second experiment the subjects were trained on series *Y* (Plate I), and tested in the manner described with series *A* (Plate II). The 24 subjects used in this group were divided into two groups of 12 each. One group was trained to move the lever to the right for stimuli of type *y*1, and to the left for stimuli of type *y*2. The other group was given just the opposite training.

In the third experiment the subjects were trained on series *Y* (Plate I), and tested with series *B* (Plate II). The 24 subjects used in this group were divided into two groups of 12 each and each group given the type of training opposite to that of the other group (as described above), to control for the effect of direction of movement.

Results

About ten percent of the subjects discovered that, for some of the test stimuli used, the direction of movement was indifferent, and their results were discarded. This left 12 subjects in each group as explained above, and as each subject had four sittings he reacted to each of the test stimuli four times. Pooling the results for the 12 subjects in a single group we have 48 responses in each group to each test stimulus. Although the subject usually reacted by moving the lever to the right or to the left, sometimes the subject failed to respond when the test card was presented, and sometimes the response consisted in moving the lever quickly back and forth from right to left. Only movements definitely to the left, or definitely to the right were considered in tabulating the results. The various ambiguous movements or failures to move were omitted. This means that the percentages calculated below are each based on from 40 to 48 responses.

The number of movements definitely to the left was added to the number of movements definitely to the right, and this sum was divided into the number of movements to the right, or into the number of movements to the left, thus giving the percentage of movements to the right, or the percentage of movements to the left. These data are given in the following tables. In most cases the percentage of movements to the right was calculated if a movement of the lever to the right had been required in response to stimuli of type *x*1, or type *y*1;

while the percentage of movements to the left was calculated if a movement of the lever to the left had been required in response to stimuli of type $x1$, or type $y1$. This may be termed the percentage of responses the subject made 'as if' he had been responding to stimuli of type $x1$ (in experiment 1), or to stimuli of type $y1$ (in experiment 2 or 3). With this method of tabulating results a 100 percent record in experiment 1 for any test stimulus will indicate that in all tests, and by all subjects that particular test stimulus was reacted to just as if it had belonged to stimuli of type $x1$. In experiments 2 and 3, 100 percent recorded for any test stimulus will indicate that in all tests and by all subjects that particular test stimulus was reacted to just as if it had belonged to stimuli of type $y1$. On the other hand 0 percent will indicate (in experiment 1) that in all tests and for all subjects that particular test stimulus was reacted to just as if it had belonged to stimuli of type $x2$. In experiments 2 and 3, 0 percent recorded for any test stimulus indicates that it was always reacted to just as if it belonged to stimuli of type $y2$.

Table I gives the data for the subjects trained to discriminate a 100° from a 25° angle, direction being the same (series X), and tested on angles which differed from both of the training stimuli only in size.

TABLE I

Training Stimuli: The two types were the same in direction (pointing towards the upper right hand corner) but different in size, a 100° and a 25° angle being used.

Test Stimuli: The direction was the same as that of the training stimuli, but the size was different. The sizes used were 5°, 12°, 140°, and 175°.

Tabular data give the percentages of cases in which each of the test stimuli elicited the response which was given to the 100° training stimulus.

Size of Angle in Degrees	5°	12°	140°	175°
% responses right for subjects trained to move lever right for stimuli of 100°	20	16	71	55
% responses left for subjects trained to move lever left for stimuli of 100°	2	9	77	74
Results for all subjects used in experiment 1	11	12	74	64

In the majority of cases each test stimulus elicited the response which had been given to that training stimulus which most resembled it in size. For example, in 74 percent of the

cases the 140° test stimulus elicited the response which had been given to the 100° training stimulus. In 12 percent of the cases the 12° test stimulus elicited the response which had been given to the 100° training stimulus. This means that in 88 percent of the cases the 12° test stimulus elicited the response which had been given to the 25° training stimulus.

The frequency with which a test stimulus elicited the response which had been given to the training stimulus most like it in size, decreased slightly with the increase of the size difference. For example, the 175° test stimulus elicited the response which had been given to the 100° training stimulus less frequently than did the 140° test stimulus.

In this experiment none of the test stimuli were identical with either of the training stimuli; and we have no data as to the strength of the habit developed in the training series as a norm of comparison.

Table II gives the data for the subjects trained to discriminate direction of angle, size being the same (series *Y*), and tested on angles which were like one or the other of the training stimuli in direction, but which exhibited varying degrees of difference in size.

In the majority of cases the angles pointing down in the test series elicited the response which had been given to the angles of the training series which pointed down; likewise in a majority of the cases the angles pointing left in the test series elicited the response which had been given to the angles of the training series which pointed left. These statements are based on the fact that most of the percentage records given exceed fifty.

The frequency with which a test stimulus pointing in a given direction elicited the response which had been given to the stimulus of the training series pointing in the same direction, decreased as the size difference increased.

With the larger differences in size, the percentages approximated fifty, as in the case of the 135° angle with vertex down, or the 175° angle with vertex left. In these cases the subjects apparently reacted on the basis of chance.

There were two rather marked exceptions to the rule of

TABLE II

Training Stimuli: The two types were the same in size (45°) but different in direction of angle—pointing down versus pointing left.

Test Stimuli: The direction was the same as that of one or the other of the two types of training stimuli, but the size was different. The sizes used were 5°, 45°, 90°, 115°, 135°, 155°, and 175° for each of the two directions. Results for the two directions are tabulated separately, as follows:

Test Stimuli with Vertex Down

Tabular data give the percentages of cases in which each of the test stimuli elicited the response which was given to the training stimuli with vertex down.

Size of Angle in Degrees	5°	45°	90°	115°	135°	155°	175°
% responses left for subjects trained to move lever left for stimuli with vertex down	79	100	78	61	57	*30*	*40*
% responses right for subjects trained to move lever right for stimuli with vertex down	76	100	64	64	37	*24*	*43*
Results for all subjects used in experiments 2 and 3	78	100	71	62	47	*27*	*42*

Test Stimuli with Vertex Left

Tabular data give the percentages of cases in which each of the test stimuli elicited the response which was given to the training stimuli with vertex left.

Size of Angle in Degrees	5°	45°	90°	115°	135°	155°	175°
% responses right for subjects trained to move lever right for stimuli with vertex left	87	96	90	93	89	72	37
% responses left for subjects trained to move lever left for stimuli with vertex left	87	98	91	91	81	59	58
Results for all subjects used in experiments 2 and 3	87	97	91	92	85	66	47

reaction to the test stimuli on the basis of their directional similarity to one or the other of the two types of training stimuli. In the majority of the cases the 155° and 175° angles with vertex down elicited the response which had been given to the training stimulus with vertex left (see figures underlined in Table II). This phenomenon will be discussed more in detail later.

Table III gives the data for the subjects trained to discriminate a 100° from a 25° angle, direction of angle being the same (series *X*), and tested on 100° or 25° angles which differed from the angles of the training series in direction.

In the majority of cases each test stimulus elicited the response which had been given to the training stimulus of

TABLE III

Training Stimuli: The two types were the same in direction (towards the upper right hand corner), but different in size, a 100° and a 25° angle being used.

Test Stimuli: The size was the same as that of one or the other of the two types of training stimuli used, but the direction was different. In one case the stimuli pointed up and slightly to the left; in the other, to the right and slightly down.

Tabular data give the percentages of cases in which each of the test stimuli elicited the response which was given to the stimuli of the same size in the training series.

TEST STIMULI OF 25°

Direction of Vertex of Angle	Up and Slightly Left	Right and Slightly Down
% responses left for subjects trained to move lever left for stimuli of 25°	100	79
% responses right for subjects trained to move lever right for stimuli of 25°	76	81
Results for all subjects used in experiment 1	88	80

TEST STIMULI OF 100°

Direction of Vertex of Angle	Up and Slightly Left	Right and Slightly Down
% responses right for subjects trained to move lever right for stimuli of 100°	87	81
% responses left for subjects trained to move lever left for stimuli of 100°	93	78
Results for all subjects used in experiment 1	90	80

the same size. That is, the response given to the 100° angles was, in the majority of cases, like the response given to the 100° angles in the training series; and the response given to the 25° angles was in the majority of cases like the response given to the 25° angles in the training series.

In each case the variation in direction was $67\frac{1}{2}$ degrees from the direction of the training series, but there is no clear indication that this same variation in direction produced the same disturbance of the response in each case.

Table IV gives the data for the subjects trained to discriminate direction of angle, size being the same (series *Y*), and tested on angles of the same size (45°), but varied as to direction.

In the table the general direction of a line of figures from the center indicates the direction of the vertex of the angle for which those figures indicate the percentage of responses.

For example, the line reading 97, 98, 96, points to the left, and indicates the results for a test stimulus of 45° pointing left, *i.e.* a test stimulus identical with the training stimulus,

TABLE IV

Training Stimuli: The two types were the same in size (45°) but different in direction of angle—pointing down versus pointing left.

Test Stimuli: The size was the same as that of the training stimuli, but the direction was different. The directions used were straight up, straight down, straight right, and straight left, with two directions between adjacent ones of these, as shown in the table.

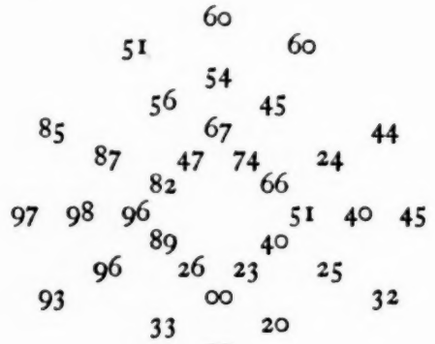
Tabular data give the percentages of cases in which each of the test stimuli elicited the response which was given to the training stimuli with vertex left.

(The general direction of the line from the center indicates the direction of the point of the angle.)

The inner circle shows the percent responses right for subjects trained to move the lever right for stimuli pointing left.

The next circle of figures shows the percent responses left for subjects trained to move lever left for stimuli pointing left.

The outer circle of figures shows the results for all subjects used in experiments 2 and 3.



and used as a norm to measure the accuracy of the habit established. The results show that in 97 percent of the cases the 45° angle pointing left elicited the response which had been given to the 45° angles pointing left in the training series. This means that in 3 percent of the cases this same test stimulus elicited the response which the subjects had been trained to give to the 45° angle pointing down. The other stimulus of the test series used as a norm is indicated in the line reading 00, 00, 00. This line points down from the center, and so indicates the results for the test stimulus with vertex down. In this case, 0 percent responses means that this stimulus was always reacted to the same as a 45° angle with vertex down in the training series.

Variations of $22\frac{1}{2}$ degrees from the stimulus pointing left are indicated by the two lines reading 85, 87, 82; and 93,

96, 89. As can be seen, test stimuli showing variations in direction of $22\frac{1}{2}$ degrees still in a majority of cases brought the response which had been given to the 45° angles pointing left. Likewise the two test stimuli varying $22\frac{1}{2}$ degrees in direction from the training stimulus pointing directly down are indicated in the lines reading 29, 33, 26; and 22, 20, 23. That is, the two test stimuli which differed $22\frac{1}{2}$ degrees in direction from the training stimulus pointing down elicited, in 22 percent and 29 percent of the cases, the response which had been given to the training stimulus with vertex left, or, in other words, in 78 percent and 71 percent of the cases these test stimuli elicited the response which had been given to the training stimuli with vertex down.

The values for variations of $67\frac{1}{2}$ degrees or more in direction are ambiguous in import. Possibly the subjects were reacting by chance.

In summarizing the preceding four tables we may designate transfer as 'unambiguous' when the percentage of responses tabulated was either 0 percent or 100 percent. Varying degrees of ambiguity of transfer would then be indicated by percentages varying from 100 percent on the one hand and from 0 percent on the other, toward 50 percent which may be termed 'complete ambiguity' of transfer, *i.e.* half the responses are in one direction, and half in the other.

When the subject has been trained to discriminate angles of two different sizes with direction of angles the same, there is transfer to novel stimuli on the basis of correspondence of size of angle. The degree of ambiguity of the transfer increases with increasing differences between the size of angle of the training and test stimuli. From the data of this experiment no confident statements can be made regarding the relationship between the degree of ambiguity of transfer, and degree of variation in direction.

When the subject has been trained to discriminate angles of two different directions with size the same, there is transfer to novel stimuli on the basis of correspondence of direction of angle. The degree of ambiguity of the transfer increases with increasing differences between the training and test stimuli in either size or direction.

The verbal reports of the subjects as to the characteristic which determined their reaction showed that one could not decide on a-priori grounds how the subject would verbalize his discrimination. In experiment 1, where the subjects were theoretically discriminating on the basis of the size of the angle, some of the subjects gave vague explanations in terms of the direction in which the arm of the angle pointed, as contrasted with the direction in which the apex of the angle pointed. In experiments 2 and 3, where the subjects were theoretically discriminating on the basis of the direction of the angle we again get a wide variety of reports, some reporting that they were reacting to the direction of the point of the angle, some to the direction of the opening. One subject distinguished between 'V's' and 'Arrows' and one between 'horizontal' and 'vertical.'

An attempt was next made to see how the subjective "similarity" of the test stimuli to the training stimuli would correlate with the percentage of responses. This method had the advantage of greater generality than the method of measuring some objective factor such as size or direction of angle, for stimuli which varied in size, or in direction, or in both could all be grouped together, and dealt with at once instead of being treated separately as in the preceding tables. So a separate experiment was conducted in which the subjects were given two sample cards of the training series, and the stack of 16 cards of the test series, with the following instructions:

> Here is a group of cards. You are to sort them with respect to their apparent similarity, as to the general appearance of the figure on the cards, to these two cards (indicating the two representative cards of the training series, which were placed at opposite corners of the table), so that finally they will be arranged in order of their similarity and dissimilarity to the two end cards. That is, as you go in this direction (indicating from left to right) the cards will become more like this card (indicating the card of the training series at the right) and less like that card; (indicating the card of the training series at the left), and conversely, as you move in the other direction the cards will become more and more like that card, and less and less like this card.

If the subjects asked just what was meant by 'similarity' they were told to interpret the word as they wished.

The subject then sorted the cards and afterwards was

asked to indicate a point which he thought was the neutral point where the cards were no more like one extreme than like the other extreme. His judgments were then recorded.

The above procedure was repeated with 48 subjects, 24 of whom had already served in the earlier part of the experiment, and 24 of whom were new subjects. Each subject was given three sorting tasks.

1. To sort the cards of test series *A*, the sample extremes being 's' of type *x*1, and 's' of type *x*2 of series *X*, (see Plate I).

2. To sort the cards of test series *A*, the sample extremes being "*s*" of *y*1, and "*s*" of *y*2, of series *Y* (see Plate I).

3. To sort the cards of test series B, the sample extremes being '*s*' of *y*1, and '*s*' of *y*2, of training series *Y* (see Plate I).

This gave three series of judgments of similarity made by 48 judges, which may be compared with the responses which were obtained in the earlier experiments, to see what relation there is between subjective judgments of similarity and the subject's reactions under the conditions of this experiment.

Although the subjects complained that they did not know what to do with such a collection of stimuli, and although the results for the different subjects gave a very wide scatter, yet it was found that when these results were pooled, the rank order of judged similarity and the rank order as to percentages of responses gave a correlation of .89 in experiment 1, .83 in experiment 2, and .93 in experiment 3. This certainly indicates a positive relationship between order of judged similarity and percentages of responses, although the correlation is probably spuriously high, since the probable error of a correlation determined from 16 cases is large, and the figures used as stimuli were selected because they showed marked differences, which would make the observed correlation still higher. The correlation can be interpreted as indicating that with the present set-up, some of the factors which affect the sorting of the cards with respect to similarity also affect the ambiguity of transfer.

The definite inverse relationship between ambiguity of transfer and similarity of test stimulus to the corresponding training stimulus make it rather certain that the similarity

of the test figure to the corresponding training figure is the principle factor influencing the subject's response to the test stimuli. However, it is well to investigate several other possible factors to be certain that this apparent influence of similarity is not in reality due to some extraneous uncontrolled factors in the experiment.

a. Chance Variation.—In experiments on animals where the same sort of imperfection in transfer has been shown it has usually been ignored or else dismissed with the statement that probably it represents some accidental variation in performance.[5] This hypothesis will not apply here. The correlations of .83 to .93 between order of percentages of responses, and order of judged similarity cannot be explained as due to any chance influence.

b. Direction of Movement.—This source of error was controlled (as explained above) by giving one half of the subjects a training just opposite to that of the other half. An inspection of the tables shows that in general when the training was reversed, the response of the subject was also reversed.

c. Type of Discrimination.—It is possible that the particular type of discrimination used is responsible for the results obtained. This factor was somewhat ruled out by having two different sorts of discrimination made; one on the basis of size of angle, and the other on the basis of direction of angle. The same results were obtained with each type of discrimination. Other work shows that the same sort of result holds for nonsense syllables, meaningful words, and complex figures.[6] It would be desirable to test still other types of discrimination. I have evidence from work done on rats with size discrimination, that the same sort of result is obtained there also. It seems safe to conclude that the particular type of discrimination made is not a dominant factor.

d. Order in Which Test Stimuli Were Presented.—The sixteen test stimuli used in any one experiment were presented to each subject in the same order at each sitting. It may be

[5] Helson, Perkins, Washburn.

[6] Yum.

that this order influenced the percentage of times the subject would respond to a given test stimulus by a movement to the right or to the left. To determine the extent of the influence of this factor the correlation between order of the test stimuli with respect to percentages of responses, and order of presentation was computed, and found to be .007 for experiment 1; − .13 for experiment 2; and − .26 for experiment 3. This shows that the order of the test stimuli relative to each other was not a dominant factor in the determination of the subject's response, although it was probably a slight factor in some cases.

e. Preceding Movements.—It may be that the series of movements made in response to the ten members of the training series which preceded any test stimulus predisposed the subject to make the next movement in a particular direction regardless of what figure was presented. No special control was run for this factor alone, but we may pool the results from the three series of tests, and compare the percentages of responses in each of the cases where the preceding series of ten movements had been identical. The average percentage of responses following any particular arrangement of stimuli may be taken as the best guess as to the effect of that order of movement. Then this average may be correlated with each of the individual members that go to make up that average, to give a rough measure of how much scatter there is about the central tendency. This correlation was found to be .53. This shows that the direction of the preceding ten movements was a slight factor, but certainly was not dominant over the actual character of the figure shown.

The evidence then shows that these five factors just considered probably had a slight influence on the response in some cases; but the primary factor was the similarity of the test to the training figures.

Discussion

The test stimuli used in the present experiment cannot be satisfactorily divided into two classes; those which are 'equivalent' to the stimuli of the training series (as shown by

the fact that the response to them was very similar to that given to the corresponding training stimulus), and those which are 'non-equivalent' to the stimuli of the training series (as shown by the fact that the response to them is a chance response). Rather the stimuli used fall into a series showing all degrees of change in percentages of responses from almost 100 percent like the response given to one or the other type of training stimulus to a practically chance or 50 percent response. This fact has been demonstrated now by at least three experiments, the present one, Yum's,[7] and an experiment on animals (now in press in the *Journal of Genetic Psychology*) by the present author. In spite of marked differences in the experimental technique of these three experiments the fundamental conclusions are the same. This would seem to indicate that this relationship between similarity of test to training stimuli, and degree of ambiguity of transfer is a fundamental fact of learning, and not just a result which is due to the peculiarities of one experimental set-up.

This fact is of interest in connection with theories of learning. Dr. Lashley has reviewed the evidence showing that an adequate theory of learning must explain not only the formation of the specific $S \rightarrow R$ connection involved in learning, but also the fact that the formation of one $S \rightarrow R$ connection establishes not only that connection, but also

[7] It might be well to present here a brief summary of the principal ways in which the set-up of the present experiment differs from that of Yum's which is the only other work known to the present writer on direct transfer of the response to a new stimulus (as distinguished from ability to learn), in adult human subjects.

1. Yum's work was done with a paired associates technique using verbal responses; in this experiment a trial and error training technique was adopted using a manual response.

2. In his work the subjects knew when they were being trained, and when they were being tested; in this experiment the subjects were ignorant of this change from the training to the test situation.

3. In his work transfer was tested 24 hours after learning, in this experiment transfer was tested right in the learning situation.

4. In his work similarity was measured either by the identity of certain spatially separable elements (as the number of identical letters in two words) or by the pooled opinions of judges; in this experiment similarity was measured by certain objectively definable relationships (such as the size or the direction of the angle), which involved no identity of certain spatially separable parts of the stimuli and also by the pooled estimates of judges.

many others. It now seems clear that this problem cannot be taken care of in terms of simple 'equivalence' of some stimuli to those of the training series, and 'non-equivalence' of other stimuli to those of the training series. The problem is much more complex than this. Any theory of learning to be adequate must explain the fact that there are varying degrees of probability that the old response will be elicited by the new stimuli, and these probabilities may be any value between 0 percent and 100 percent.

Summary

1. There is transfer on the basis of size of the angle when the subject has been trained to make a size discrimination, and on the basis of direction of angle when the subject has been trained to discriminate direction of angle.

2. There are varying degrees of ambiguity of transfer on the basis of size or of direction of the angle.

3. Several explanations of the varying degrees of ambiguity of transfer were considered. The hypotheses of chance variation, direction of movement, the type of discrimination made, the order of the test stimuli, or the direction of the immediately preceding series of movements were found to be inadequate. The general similarity of the figure on the card in the test series to the figures used in the training series was found to be an adequate explanation of the results.

4. The technique of using interpolated test trials with punishment removed can be used on human subjects without their becoming aware of the fact.

5. The nature and variety of the subject's verbal reports as to the basis of his discrimination cannot be determined from an inspection of the stimuli used in training, and the responses made by the subjects.

6. The test stimuli used cannot be divided into two classes, equivalent and non-equivalent; the data must be treated in terms of degree of resemblance. This fact is important in connection with theories of learning.

Bibliography

Gayton, A. H., The discrimination of relative and absolute stimuli by albino rats, *Journal of Comparative Psychology*, 1927, 7, 93–105.

Helson, Harry, Insight in the white rat, Journal of Experimental Psychology, 1927, 10, 378.

KLÜVER, HEINRICH, The equivalence of stimuli in the behavior of monkeys, *Journal of Genetic Psychology*, 1931, **39**, 3–27.

LASHLEY, K. S., Basic neural mechanisms in behavior, *Psychological Review*, 1930, **37**, 1–24.

LASHLEY, K. S., The mechanism of vision: I. A method for rapid analysis of pattern-vision in the rat, *Journal of Genetic Psychology*, 1930, **37**, 453–460.

PAVLOV, I. P., *Conditioned Reflexes*, London: Oxford University, 1927.

PERKINS, F. THEODORE, AND WHEELER, RAYMOND HOLDER, Configurational learning in the goldfish, *Comparative Psychology Monographs*, 1930, **7**, No. 1.

RUEDIGER, W. C., Indirect improvement of mental function through ideals, *Educational Review*, 1919, **36**, 364–371.

THORNDIKE, E. L., Mental discipline in high-school studies, *Journal of Educational Psychology*, 1924, **15**, 1–22, 83–98.

VOLKMANN, A. W., Über den Einfluss der Uebung auf das Erkennen Räumlicher Distanzen, *Berichte über die Verhandlungen der königlich sächsischen Gesellschaft der Wissenschaften zu Leipsig. Mathematisch-physische Classe.* Leipsig: Hirzel, 1858, **10**, 38–69.

WASHBURN, M. F., AND ABBOTT, E., Experiment on the brightness value of red for the light adapted eye of the rabbit, *Journal of Animal Behavior*, 1912, **2**, 145–180.

WATSON, J. B., AND RAYNER, ROSALIE, Conditioned emotional reactions, JOURNAL OF EXPERIMENTAL PSYCHOLOGY, 1920, **3**, 1–14.

WOODROW, H., The effect of types of training upon transference, *Journal of Educational Psychology*, 1927, **18**, 159–172.

WYLIE, H. H., An experimental study of transfer of response in the white rat, *Behavior Monographs*, No. 3, 1919.

YUM, KWANG SUP, An experimental test of the law of assimilation, JOURNAL OF EXPERIMENTAL PSYCHOLOGY, 1931, **14**, 68–82.

(Manuscript received June 29, 1931)

LEARNING DURING DIRECTED ATTENTION

BY WM. C. F. KRUEGER

College of the City of Detroit

Statement of Problems.—1. Is specifically directed attention during learning more economical than nondirected effort?

2. Does the relative influence of directed attention upon the rate of learning vary with the section of the unit to which specific attention is given?

3. When a unit to be learned contains easy and difficult materials, will it be more economical to direct the attention at first to the easy or to the difficult items?

EXPERIMENT I

Materials.—The subjects were required to memorize lists of paired associates. Each list contained 12 pairs of logically unrelated nouns. The first member of each pair was used as the stimulus-word for the recall which followed immediately after the presentation of the series. The serial order of the stimulus-words during recall was different from that of the order presented during learning. This would prevent the subjects from memorizing the response-words in the serial order of their presentation.

In order to facilitate the direction of attention the pairs of words were underlined in black or in red. One-half of the lists had the first three and the last three pairs underscored in red while the middle six pairs of associates were underlined in black. For the other half of the lists the order of marking was reversed; the middle six pairs were underscored in red while the rest had the black lines under them.

The following method was used to equate the difficulty of the various lists of paired associates. Eight sets of 24 words were selected. Each group of 24 words was then arranged to form 12 pairs of associates. From each of the 8 sets of 24 words 36 different pairs of associates were formed. These were arranged into three lists. Thus there was a total of 24 lists, half of which were used during the practice period and the other half employed for the experiment proper.

Subjects.—The 72 subjects were undergraduate college students who had little or no previous training in psychological laboratory methods. The three groups of 24 subjects seemed fairly well equated in the estimation of the experimenter.

A list of twelve pairs of nouns was shown by means of a Chicago memory drum at the rate of 2 seconds per pair of words. Only one pair of words was visible at a time. Immediately after having seen the list, the subject was shown the first member of a pair of associates, and he was to recall the correct correlate. Five seconds were allowed within which to make a written recall; then another stimulus-word was presented. Only the correct answers were counted when they were started or finished within the five-second period. Learning and recall were alternated until one complete, errorless

recall was made. No rest-periods were interpolated between the trials. The subjects were instructed repeatedly not to rehearse or anticipate any other pair of associates than that which was visible during the 2-second interval for exposure.

All subjects mastered four practice lists before starting to memorize lists for the experiment proper. Only one set, or list, was learned on any one day.

One group of 24 subjects (control group) was told to learn the series as quickly as possible. No direction of attention during learning was given. Each subject attempted to master the material as rapidly as he could by his own methods. The other groups were instructed definitely in respect to which of the sections special attention should be given. By special attention was meant that the subject should attempt to learn the assigned section or sections first, and then learn the rest of the series not yet memorized. For example, for two of the four lists to be memorized for the experimental data, the subject tried to learn the first three and the last three pairs first, and after that the rest of the list not yet mastered. For the other two lists the order was reversed in that the subject attempted to master the middle six pairs first and later the rest of the series. By means of the black or red lines under the pairs of nouns it was possible to direct attention rather accurately. For each condition of learning one list had the middle section underlined in black and another list underscored in red. The same was true in respect to the first three and the last three items. Thus by carefully counterbalancing the *order* of directing attention and counterbalancing the *order* of underscoring the words no one condition of learning had any advantage over any other condition other than in the matter of directed attention itself. The same lists were used by the control group to which no direction of attention was given, and by the groups which received specific instructions.

Results

The results are compiled in two sets of data. One set, Table I, indicates the respective number of trials necessary to memorize the lists under the various conditions of learning. The other set of data, Table II, shows the frequency of correct recalls for each serial position within the list during the first four trials. In a few instances the subject succeeded in memorizing the series in four trials, therefore it seemed best to the writer to limit the tabulations in the second set of data to the first four trials.

Table I shows that the direction of attention to the middle section of a series had no advantage over directing the attention to the initial and final members of the list. The former method required 6.08 trials while for the latter method an average of 6.17 trials were necessary. The difference between the two means is too small to be considered of significance. When no special direction of attention was given, an average of 6.48 trials was required for the same lists. Therefore, the material was learned somewhat more quickly

TABLE I

AVERAGE NUMBER OF TRIALS REQUIRED FOR COMPLETE MEMORIZATION

Method of Directed Attention	Av. No. Trials	S.D.	P.E.	$\frac{D}{\text{P.E. (diff.)}}$
No specific direction given.....	6.48	1.48	.14	When compared with average for non-directed attention
Attention directed to center.....	6.08	1.44	.14	2.00
Attention directed to beginning and end.....	6.17	1.22	.12	1.61

under conditions of directed attention than when the subjects were left to their own methods. The resultant difference is small and the probability of a true difference is approximately 86–91 chances out of a 100 as determined by $\frac{D}{\text{P.E. diff.}}$.[1]

On the basis of introspective reports given by the subjects (although much reliability cannot be given to the reports because the subjects were untrained in giving introspections) it appeared that when the subjects were not told specifically how to direct the attention, a shifting in method of learning would continue at times. When specific direction was given, there seemed to be a definite plan of procedure and the subjects settled down to the learning of the material rather than to the finding of more economical means of learning.

In Table II are listed the total number of correct responses for the various serial positions and for the three methods of learning during the first four trials. Having counterbalanced the known variables such as the relative difficulty of the various lists and the items within the lists, further practice effects, learning ability of the subjects, etc., it seems to the writer that any differences found must be due to differences in the given direction of attention. It is readily seen that

(1) When attention was directed to the first three and the last three items of the series, these items were learned more readily than the middle section of the series;

(2) When attention was directed to the middle six items of the list, these were memorized more readily than the first and the last three items;

[1] William S. Foster, Experiments in Psychology, pp. 77–9. New York: Henry Holt & Co., 1923.

TABLE II

FREQUENCY OF CORRECT RESPONSES FOR ALL SERIAL POSITIONS DURING FIRST FOUR TRIALS

Trial	Direction of Attention	Serial Position within the List												Total Recalled
		1	2	3	4	5	6	7	8	9	10	11	12	
First	None	23	16	15	11	10	8	10	9	9	12	15	18	156
First	Center	10	13	13	16	20	17	19	17	23	8	7	6	167
First	Ends	23	17	20	9	10	8	9	8	12	24	18	16	174
Second	None	31	25	20	20	19	20	18	18	19	22	25	25	254
Second	Center	16	19	19	30	25	30	25	27	29	10	14	14	258
Second	Ends	27	27	30	15	19	14	16	11	19	29	33	21	261
Third	None	34	30	28	25	22	24	24	24	24	31	33	34	327
Third	Center	20	21	23	35	30	36	31	37	35	15	25	25	333
Third	Ends	30	32	32	20	22	20	22	22	25	35	36	34	330
Fourth	None	36	34	32	31	30	30	29	28	30	36	36	38	389
Fourth	Center	26	27	26	43	42	41	36	39	39	26	29	29	403
Fourth	Ends	39	40	37	25	25	26	27	29	30	39	41	39	397

(3) The total number of items learned by either of the above methods was approximately the same, both in terms of number of trials required for complete learning and in terms of the actual number of items recalled per trial;

(4) When the subjects were left to their own devices of learning, the usual order of learning took place; namely, the first and the last items were learned sooner, and the nearer the item was to the center of the list, the later it was mastered;

(5) When no specific direction of attention was given, the subjects memorized fewer items for the same number of trials than when specific direction was furnished.

EXPERIMENT II

Subjects and Materials.—Three groups, each of 24 college undergraduates, memorized the serial sequence of 12 geometric nonsense figures. The experimenter tried to equate the groups on the basis of initial learning ability shown during the practice periods and class standing in terms of class grades. The designs were drawn with India ink on 3″ x 5″ cards, one figure to a card. The area contained within the lines of each design was the same for all figures. Eight sets of cards were used, four during the practice period and the

other four series, twelve cards to each set, while obtaining the data to be presented.

Experimental Procedure.—A set of cards was shown to the subject at the rate of two seconds per card, only one card being visible at a time during the learning period. Immediately after having observed the twelve designs in succession, the subject was given a pack of twelve cards containing the same twelve designs as were shown him at first. He had been instructed previously to rearrange the cards, so that the newly arranged order of the designs would be the same as that observed while learning. Two minutes were allowed for sorting the cards. After the two-minute interval the subject was shown the original sequence again in the same manner as before. Then he was handed another set of the same twelve designs and again instructed as before, to arrange these cards in proper sequence. 'Learning' (observing the original set in a definite sequence) and 'recall' (rearranging the twelve designs in an order to correspond with the order of the observed succession) alternated for five times.

The sequence of the designs was changed for each subject. This prevented the possibility of any one design or group of designs remaining in any favored serial position. A record was kept for each trial. Only the cards which were placed correctly, were scored. For example, the design seen as the third item during learning, had to be placed as the third item in the recall order.

Only one series was learned on any one day.

Three methods of learning were used.

Group *A.*—The subject was instructed to attempt to learn the order of the first four and the last four designs first, and later to learn the order of the middle four designs.

Group *B.*—The subject was to direct his attention especially to the middle four designs and learn their sequence first, and then to memorize the order of the first four and the last four members of the series.

Group *C.*—The subject was told to learn the entire sequence of the twelve cards as quickly as possible. No direction of attention was given and the subject left to his own devices for learning.

For the recall all subjects were asked to place as many of the twelve cards in their respective serial position as they could, irrespective of the direction of attention given and the method of learning.

Results

The results are found in Table III. Again it appeared that

(1) The middle four designs were learned more quickly when attention was directed to them than when attention was given to the first four and the last four items;

(2) The first four and the last four figures were memorized more readily when attention was focused upon them than when attention was centered upon the middle four items;

TABLE III

FREQUENCY OF CORRECT RESPONSES FOR ALL SERIAL POSITIONS DURING THE FIRST FIVE TRIALS

Trial	Direction of Att'n	Serial Position within the List												Total Recalled
		1	2	3	4	5	6	7	8	9	10	11	12	
First	None	58	40	31	25	21	19	15	12	15	20	29	57	332
First	Center	46	35	32	20	48	27	31	23	20	16	28	36	353
First	Ends	65	39	32	31	20	18	19	17	12	16	25	51	345
Second	None	131	99	79	54	40	44	48	40	51	53	81	116	836
Second	Center	110	86	55	52	101	64	68	65	43	32	65	98	839
Second	Ends	145	95	82	74	43	41	40	49	29	44	66	122	830
Third	None	220	163	146	130	93	83	90	85	84	101	136	205	1536
Third	Center	181	144	111	98	184	122	128	130	85	58	109	155	1605
Third	Ends	238	180	149	128	86	77	71	80	84	98	140	210	1541
Fourth	None	279	218	192	170	146	136	139	144	161	167	191	273	2216
Fourth	Center	241	206	168	159	259	183	199	196	144	115	165	243	2245
Fourth	Ends	319	250	209	199	141	130	118	139	118	148	187	287	2278
Fifth	None	394	323	276	242	206	198	201	189	180	224	287	356	3086
Fifth	Center	346	279	231	220	330	243	272	264	209	174	230	326	3134
Fifth	Ends	411	337	294	270	206	199	167	181	200	215	270	374	3124

(3) When the subjects were dependent upon their own methods of learning, the usual results were obtained; learning proceeded from the ends toward the center;

(4) The total number of items learned per trial for each of the two methods of directed attention is approximately the same;

(5) When the scores for methods of directed attention are compared with the scores for the method of nondirected attention, a larger number of items was learned by both methods of directed attention for each particular number of trials.

EXPERIMENT III

Subjects and Materials.—A group of 32 undergraduate students participated. Each list to be memorized consisted of five pairs of three-letter nonsense syllables and five pairs of monosyllabic nouns. It was attempted to select and group the material within each list in such a way as to avoid logical relationships or obvious connections between any two members of each pair or any two pairs within the list. Four methods of grouping were used; namely,

(*A*) 5 pairs of words followed by 5 pairs of syllables;
(*B*) 5 pairs of syllables followed by 5 pairs of words;
(*C*) pairs of nouns alternating with pairs of syllables, the series beginning with nouns;
(*D*) pairs of syllables alternating with pairs of nouns, the series beginning with syllables.

Each of twelve sets of material were arranged in accordance with the above outline.

Experimental Procedures.—The method of presenting the lists was the same as in Experiment I. All subjects practiced on at least one different list for each of the four methods of presentation before any data was collected for the investigation. By carefully counterbalancing the order of the four methods of showing the material to the learner, the effects of continued practice were compensated for all the conditions of learning. Obviously, a different set of items was used for each of the four methods for each subject.

Before starting to learn a new list the subject was instructed (1) to attempt to learn the words first and after that to complete the learning of the series, or (2) to attempt to master the syllables first and then complete the memorization of the list. During recall the subject was asked to recall both the nouns and the syllables, irrespective of the method of presentation and the direction of attention. Each subject was repeatedly requested to study only the items visible during the 2-second period and he was instructed not to rehearse nor anticipate any other pair of items.

Results

The results are based upon 32 scores for each of the four methods of presentation and the two ways of direction of attention. From Table IV it can be seen that

(1) The average number of trials required for learning a list does not vary significantly for all the conditions of learning; *i.e.*, it makes but little difference with respect to the total of effort involved, whether such effort is directed to either the easier or the more difficult material.

(2) Attention directed at first to the difficult material

(nonsense syllables) required slightly fewer trials than when attention was directed at first to the easier items (nouns).

(3) When difficult and easy material alternated, the learning was slightly faster than for the conditions in which the nouns and the syllables appeared successively in groups of five nouns or five syllables.

TABLE IV

TRIALS REQUIRED FOR LEARNING

Grouping of Material	Attention on Nouns			Attention on Syllables		
	Av. No. Trials	S.D.	P.E.	Av. No. Trials	S.D.	P.E.
Method *A*...	8.53	2.11	.25	8.39	1.89	.22
Method *B*...	8.58	2.09	.25	8.32	1.94	.23
Method *C*...	8.27	2.01	.24	8.04	1.86	.22
Method *D*...	8.16	1.79	.21	7.99	1.72	.20

DISCUSSION

Experiments I and II were conducted to study the effects of directed attention to various sections of the materials to be learned as a means for more economical learning. An article published by Rich [2] concludes that

"special attention to the middle of a series, when it can be obtained, is of advantage in learning serial material, but not in learning material of nonserial nature."

This statement is based upon the data obtained from one 'trained' subject. The results from two 'untrained' subjects were rejected on the basis of alleged incorrect procedure and incompetent report on the part of these two subjects. No definite numerical data were given. However, it was stated that the results reported by the 'untrained' subjects were not in agreement with those of the 'trained' subject.

In Experiment I the subjects mastered nonserial material (paired associates) while in Experiment II the material was of a serial nature. With both types of material, it seems evident that irrespective of where attention was directed, the total amount learned for each method of directed attention was approximately the same, both (*a*) in terms of the total

[2] Rich, G. J., Directed attention and learning, *Journal of Educational Psychology*, 1917, vol. 8, 239–240.

number of trials required for complete mastery, and (*b*) in terms of the total number of items recalled during each respective trial or recall. When no specific direction of attention was given and the subjects were left to their own methods of learning and directing of attention, the learning was slightly slower. Aided by information gained from the introspective reports, the following explanations may be offered.

1. Whenever a specifically directed attention prevailed, the subjects usually would settle down at once to a specific method of mastering the material in accordance with the instructions given.

2. Whenever the subjects were left to their own devices, some would keep on shifting in methods of learning, hoping to find a more efficient method for memorizing.

3. Directed attention usually resulted in learning by some sort of 'part method' while many of the subjects in non-directed attention would learn by the 'whole' method.

4. Learning without specifically directed attention usually would result in giving greater attention to the beginning of the series, and the last members seemed to stand out more clearly because no other items followed to mar the clearness of perception and retention of these last items.

5. When the subjects learned by the methods of directed attention, they found that the number of items to which they attended were rather frequently within the span of perception.

6. The determination of the subjects to master the material and the intensity of impression were greatest during the perception of the initial items, and particularly during the first presentation of the list. The subjects began to get 'mixed-up' after the fourth or fifth item in the series had been presented, and with each successive new item the clearness of the more immediately preceding items became dimmer. The last items of the series stood out more clearly because no new items came to make for confusion.

7. Since the time-interval between the presentation of the last item in the learning procedure and the presentation of the first item in the recall series was exactly the same as

between any two items during the learning period, it cannot be said that the subjects had the opportunity to repeat the last item more frequently than the others, and thereby explain recency or finality on the basis of frequency of stimulation.

8. The speed of mastering a series, or parts of the series, may be determined by the clearness of perception, the determination or selection on the part of the learner in terms of directed attention, the span of perception, and the inhibitory or confusing influence of one item upon another.

Experiment III was performed to observe the effect of directed attention to materials of varying difficulty within the same unit to be learned. A preliminary study showed that it required an average of 6.03 trials to memorize a list of 12 pairs of nouns, while it took 12.64 trials to master the same number of nonsense syllables. The same syllables and nouns were used in Experiment III. Therefore, the nonsense syllables clearly represent the more difficult material.

Table IV shows that when attention was directed at first to the difficult material (nonsense syllables), the average number of trials necessary to learn a list is slightly smaller than when attention was given at first to the easy material (nouns). However the difference between the two means is not to be considered as very significant when determined by the interpretation [3] of an application of the formula $\frac{D}{\text{P.E. diff.}}$ to calculate the statistical reliability of the difference between the two means.

When the material was presented so that a pair of nouns alternated with a pair of nonsense syllables, the average number of trials required for learning was slightly less than when the same material was presented so that the five pairs of nouns and the five pairs of nonsense syllables were grouped (Methods *C* and *D*).

This may be explained in part by some of the introspective reports given. When the nouns and the syllables alternated, a sort of 'rest period' occurred whenever the intensity of attention was decreased with each alternating pair of items to

[3] William S. Foster, *op. cit.*

which no particular or special attention was to be given. This was more noticeable for the arrangement which required attention upon the syllables. These 'rest periods' could not be interpolated when the subject was required to give special attention to the five consecutive pairs of associates (Methods *C* and *D*). Furthermore, in spite of the repeated instructions not to rehearse nor to anticipate any pair of associates other than the one visible at the time, some subjects reported that they could not prevent thinking of or even repeating silently some of the other pairs now and then. This was particularly true during the early part of the practice period; and, with continued practice, these rehearsals became consistently fewer.

Conclusions

On the basis of the experimental data obtained one may conclude:

1. Directed attention brings about more economical learning than nondirected effort. This holds true for serial and nonserial material.

2. When the attention given is constant, the total amounts learned are the same, irrespective of whether this effort is directed to the beginning, the center, or the final sections of the unit which is to be memorized.

3. Directed attention facilitates the learning of the material attended to, irrespective of its serial position within the unit, during the learning process.

4. When materials of unequal difficulty appear within the same unit to be mastered, the total number of trials required to memorize the unit is approximately the same whether attention is given at first to the more difficult or the easier sections of the unit.

5. When difficult and easy material alternate frequently within the same unit, the entire unit is memorized slightly faster than when the same material is grouped into successively easy and successively difficult materials.

(Manuscript received May 28, 1931)

A STUDY OF THE EYE-MOVEMENTS OF STUTTERERS DURING ORAL READING [1]

BY HERBERT H. JASPER
University of Iowa

AND

ELWOOD MURRAY
University of Denver

This study is particularly concerned with the comparison of normal speakers and stutterers as to the nature of their respective eye-movements during oral reading, and with the eye-movements of stutterers during stuttering spasms. Since the behavior of the eyes is such a sensitive indicator of many nervous and mental diseases it was thought that the behavior of the stutterers' eyes during speech might shed some light on the nature of the nervous disorder in stuttering.

As far as the present authors are aware, there are on record no investigations of the eye-movements of stutterers during oral reading. Most of the data available from the photographic study of eye-movements pertain to the silent reading and fixation movements of normal speakers. Huey (**6**) and Schmidt (**9**) have summarized the work of earlier experiments in this field, while W. S. Gray (**5**) and Tinker (**10, 11**) have given comprehensive summaries of the more recent studies. The several studies concerning eye-movements have all been made in connection with studies of silent reading. Among these researches have been the studies of Schmidt (**9**) and C. T. Gray (**4**) with which at a later point we shall compare the results of our study. Comparisons will also be made with the eye-movements of stutterers during silent reading, a subject recently investigated by Murray (**8**).

[1] This paper is a preliminary report of the investigation of the eye-movements of stutterers being carried on at the psychological laboratories of the University of Iowa as a part of the large program of research on the nature of stuttering under the direction of Dr. Lee Edward Travis.

Photographic studies of eye-movements in the insane have been made by Diefendorf and Dodge (**2**) which will likewise be given consideration in the interpretation of the results of the present study.

Our subjects were fifteen stutterers with ages ranging from 10 to 29 years and grades ranging from the fifth grade to graduate standing in college. Mild, moderate, and severe stutterers were about evenly represented among the subjects.

The reading material consisted of selections from text-books in the social sciences. Four different selections were made of material varying in difficulty so that it could be adapted to the grade levels of the different subjects.

The apparatus for the photography of eye-movements was a somewhat simplified and modified form of the cameras that have become more or less standardized by Dodge (**3**), Dearborn (**1**), Schmidt (**9**), Miles (**7**), Tinker (**10**), and others. The source of light was a 100-watt concentrated filament nitrogen filled lamp. The corneal reflection of both eyes was photographed on Eastman standard motion picture film through two lenses of the proper strength to magnify the movement of the corneal reflection approximately 4.5 times. The film was moved at a continuous and constant speed by an electric motor. Time values were indicated on it by means of mercury filled neon light flashing twice a second. The flash was timed by a stroboscopically regulated phonograph motor upon which was attached a cam and lever. The lever was attached to a rocker switch which made and broke contact in a mercury cup.

The apparatus herein described is designed for only the horizontal movements of both eyes. We were particularly interested in binocular coördination as well as monocular movement so that we did not deem it advisable to sacrifice the horizontal movement coördination of one eye in order to obtain vertical movements. On our records these latter movements are indicated by the piling up and thinning out caused by the light rays moving either with or opposed to the direction of the film.

The data presented were treated both quantitatively and

qualitatively. The quantitative treatment involved ascertaining the average number of fixations per line, the average number of regressive movements per line, the average duration of fixations, and the average perception time per line. The qualitative study consisted of an analysis of the records for typical movement patterns, extraneous movements and other phenomena not amenable to quantitative treatment. The quantitative measurements for each of the fifteen stutterers are summarized in Table I.

TABLE I

QUANTITATIVE DATA ON EYE-MOVEMENTS OF STUTTERERS DURING OVERT READING

Subjects	Average Number of Fixations per Line	Average Number Regressions per Line	Average Duration of Fixation, in Sigma	Average Fixation Time per Line, in Sigma
Ste.	14.85	4.03	226	3356
Dow.	11.80	2.64	259	3056
Gar.	8.25	.71	245	2021
McK.	10.71	3.13	266	2848
Rie.	13.91	3.73	248	3449
Gus.	8.23	2.08	239	1966
Fis.	13.36	4.09	327	4368
Cot.	10.00	3.20	378	3780
Cas.	10.03	2.00	242	2427
Vaw.	10.66	5.20	601	6406
Joh.	10.00	1.80	298	2980
Van.	12.00	3.50	517	6216
Mao.	10.55	2.66	256	2700
Bul.	11.50	2.50	235	2702
Gri.	10.90	3.40	202	2201
15.	Av. 11.11 S.D. 1.79 S.D. Av. .46	Av. 2.97 S.D. 1.06 S.D. Av. .27	Av. 302 S.D. 109 S.D. Av. 28	Av. 3365 S.D. 1337 S.D. Av. 345

Comparable groups from the studies of C. T. Gray and W. A. Schmidt are to be had as normals for a comparison of the results obtained on our fifteen stutterers. The groups selected from the former studies are comparable in age range and in school standing with our group of stutterers. Conditions of reading such as length of lines, size of type, and difficulty of reading matter is sufficiently similar to allow comparisons to be made.

The average number of fixations per line for stutterers 11.11 ($S.D._{av.}$ .46), is greater than the average number of fixations per line, 9.80 ($S.D._{av.}$ .26), for 28 of Gray's pre-

sumably normal speakers (4 : *p* 100). The difference of 1.31 can be considered statistically significant considering its standard deviation of .53. Seventeen comparable subjects from Schmidt's study (9 : *p* 41) averaged 8.6 ($S.D._{av.}$ .21) fixations per line which is 2.51 ($S.D._{diff.}$ .51) less than the average number of fixations per line for our stutterers.

The stutterer's average of 2.97 ($S.D._{av.}$ .27) regressive movements per line compares with an average of 2.18 ($S.D._{av.}$ .25) and 1.88 ($S.D._{av.}$ .02) from the studies of Gray and Schmidt respectively. The differences of .79 ($S.D._{diff.}$ .37) and of 1.09 ($S.D._{diff.}$ .27) indicates a tendency to more regressions per line on the part of our stutterers.

We find no significant differences between the duration of fixation pauses in the stutterers and the normal speakers when the averages alone are taken into consideration. This can be explained by the fact that the stutterers compensate for a few very prolonged pauses by unusually short pauses.

Since fixation time per line is the product of the average number of fixations per line and the average duration of fixations, the same differences would be found in this calculation as have been mentioned above in its components.

The following table summarizes the quantitative comparisons of stutterers and normal speakers in oral reading.

TABLE II

DIFFERENCES BETWEEN NORMALS AND STUTTERERS IN EYE-MOVEMENTS DURING ORAL READING

(Differences expressed in terms of $S.D._{diff.}$)

	Gray	Schmidt
No. fixations per line	2.5 S.D.	4.9 S.D.
No. regression per line	2.1 S.D.	4.0 S.D.
Duration of fixations	1.0 S.D.	2.3 S.D.

This table indicates that the greatest difference between stutterers and normal speakers in oral reading is to be found in the larger number of fixations per line.

Murray's investigation (8) of eighteen stutterers (fifteen of which were the same subjects used in this study) makes the data available for a comparison of the differences between the eye-movements in oral and silent reading within the

same groups and between groups for both stutterers and normals. Table III presents the results from the several researches that permit comparisons to be made. In all phases of the reading process with the possible exception of the average duration of fixations, it is apparent that the stutterers have greater differences between silent reading and oral reading than have the normals.

The recording of eye-movements during stuttering spasms was the most difficult aspect of this study. The closing of the eyes and the movements of the head during the clonic and tonic blocks of the speech mechanism made it necessary for us to discard many of the stutterer's records. We finally obtained enough records of eye-movements during these blocks to be able to draw some fairly definite conclusions in regard to the behavior of the eyes during stuttering. The actual records obtained were supplemented by watching the movements of the eyes from the bright spots on the film while the records were being taken. These observations served to check the actual records and in some instances to reveal the nature of the movements during spasms to a better advantage than did the records themselves.

Indications of synchronism between the eye-movements and tonic and clonic blocks of the speech mechanism were obtained only by observing the movement of the spots on the film and listening to the speech of the reader while the record was being taken. In all records it appeared to the observer that there was a marked synchronism between vertical twitches of the eyes and clonic blocks of the speech mechanism and between prolonged fixations of the eyes and tonic blocks in the speech mechanism. The rebuilding of the apparatus [2] has made possible a more accurate indication on the record itself of the incidence of certain types of speech disturbances. The work thus far with this improved apparatus has served to corroborate the results obtained by the methods used in the present study.

Figure 1 is a typical record of subject Dow., who is a

[2] This apparatus is described in an article entitled 'The Iowa Eye-movement Camera' to appear in *Science*.

TABLE III

COMPARISON BETWEEN ORAL AND SILENT READING

	Av. No. of Fixations per Line	Av. No. of Regressions per Line	Av. Duration of Fixations (Sigma)	Av. Perception Time per Line (Sigma)
STUTTERERS:				
Oral Reading	11.11—$S.D._{av.}$ .46	2.97—$S.D._{av.}$ .27	302—$S.D._{av.}$ 28	3365
Silent Reading				
Murray	7.93—$S.D._{av.}$ .28	1.45—$S.D._{av.}$ .06	235—$S.D._{av.}$ 34	1780
NORMALS:				
Oral Reading				
Gray	9.80—$S.D._{av.}$ .26	2.18—$S.D._{av.}$ .25	262—$S.D._{av.}$ 24	2568
Schmidt	8.60—$S.D._{av.}$ .21	1.88—$S.D._{av.}$ .02	373—$S.D._{av.}$ 14	3207
Silent Reading				
Gray	7.40—$S.D._{av.}$ .43	1.30—$S.D._{av.}$ .18	258—$S.D._{av.}$ 9	1909
Schmidt	7.07—$S.D._{av.}$ .31	1.21—$S.D._{av.}$ .17	311—$S.D._{av.}$ 11	2198
DIFFERENCES BETWEEN ORAL AND SILENT READING:				
Stutterers	3.18—$S.D._{diff.}$ .53	1.52—$S.D._{diff.}$ .28	67—$S.D._{diff.}$ 44	
Normals				
Gray	2.30—$S.D._{diff.}$ .50	.88—$S.D._{diff.}$ .31	4—$S.D._{diff.}$ 26	
Normals				
Schmidt	1.53—$S.D._{diff.}$ .37	.67—$S.D._{diff.}$ .17	62—$S.D._{diff.}$ 26	

comparatively mild stutterer presenting mainly tonic blocks in stuttering. (All records are read from left to right.) The record from *A* to *B* is the horizontal movement of the two eyes in the reading of one line. The upper line is the right eye, the next line is the left eye and the base line is the record of head movements. The space included within any three of the spots along the head line represents a duration of one second. The onset of a block is represented in the eye-movements of this subject by a prolonged fixation of the word causing difficulty. However, the eyes do not always fixate the word upon which blocking occurs.

Figure 2 of subject Mas. is characteristic of the type of stutterer with few but rather prolonged tonic spasms which break into clonic spasms until the speech mechanism becomes sufficiently reorganized to form the word attempted. This record shows the end of a tonic spasm which lasted six seconds and the clonic spasms of recovery before normal reading movements begin.

Figure 3 is a very unusual record of eye-stuttering spasms of subject Rie. during attempted oral reading.

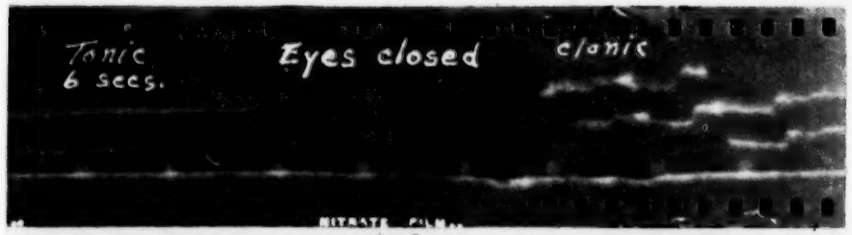
Figures 4 and 5 are characteristic of eye-stuttering spasms of subject Cot., who is a quite severe stutterer with clonic spasms usually preceding tonic spasms. Fig. 4 begins with a tonic spasm which was initiated by a series of clonic spasms not shown in the record. Particularly notable in this record is the marked head movement associated with the clonic spasms. Fig. 5 begins at the end of a series of normal reading movements and shows particularly well the blinking of the eyes and vertical twitches preceding the tonic spasm. In this

PLATE I

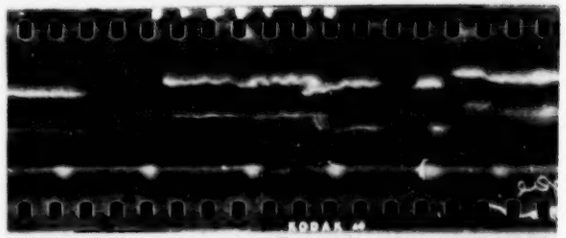
Records of eye movements of stutterers in oral reading. In each record reading from the top downward the first line represents movements of the right eye; the second line, movements of the left eye; and the third line, movements of the head. In normal readers eye-movement lines begin near the bottom of the record corresponding to the left side of a page, and by a series of rhythmical jerks proceed to the right and upwards with the last fixation on the page indicated by the two parallel lines nearest the top of the film. The white spots show vertical twitches of the eyes. The gaps in the lines are the result of closing the eyes. Superimposed upon the head line are the white spots made by the timing lamp. The space included within any three white spots represents one second.

PLATE I

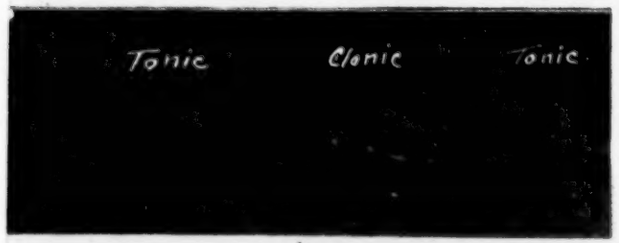
1



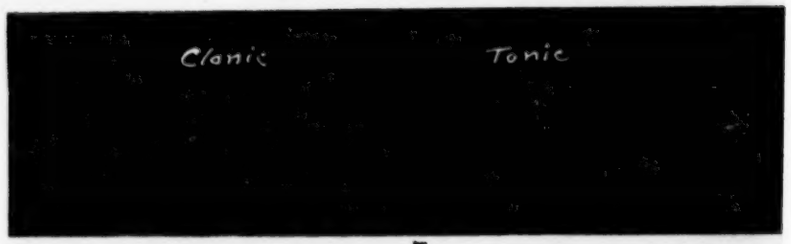
2



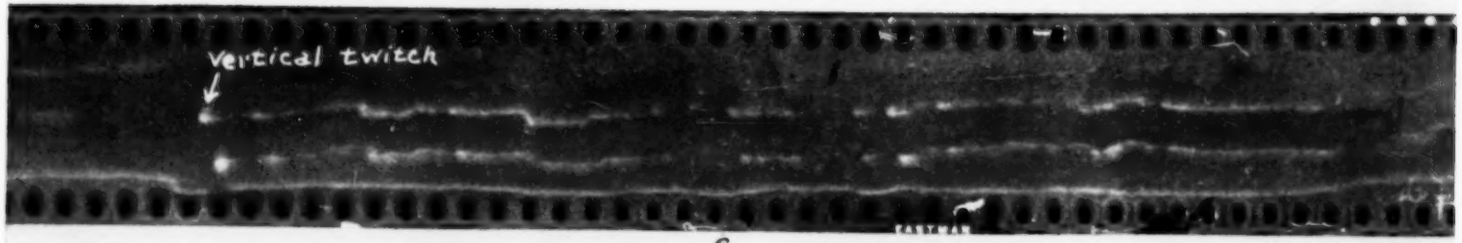
3



4



5



6

particular section of the record, the eyes were perfectly rigid in fixation during the tonic spasm, closed just before the clonic spasms began, and showed irregular fixation movements, with marked vertical twitches during the clonic spasms.

Figure 6 of subject Fis. demonstrates marked disintegration of fixation movements during a prolonged series of tonic and clonic spasms. This subject was a very severe stutterer whose eye-movement record during a spasm showed very irregular fixations, marked vertical movements, and many regressions. There were also considerable head movements and blinking of the eyes. The stuttering spasms of this subject were characterized by unvocalized 'spitting' movements of the articulatory mechanism.

From the many records obtained similar to those described above, there can be no doubt concerning a marked disturbance of eye-movements during stuttering spasms. Furthermore, there seems to be suggestive evidence for a means of differentiating types of spasms by means of the records of eye-movements during these spasms. We have observed four types of eye-movements during spasms from the records obtained so far: (1) Rigid fixation of the two eyes throughout the spasm; (2) vertical twitches of the eyes at the onset of the speech disturbance followed by a rigid fixation or slow continuous horizontal movement; (3) rigid fixation or slow continuous horizontal movement at the onset of the spasm usually followed by the closing of the eyes and then marked vertical twitches and irregular fixations until normal speech was resumed; and (4) general chaotic disturbance such as that in Fig. 6, where the eyes alternate between vertical twitches and rapid aimless horizontal movements. We find that each of the stutterers we have studied thus far may be fairly definitely classified into one of the above classifications and that careful observation of the disorders in the speech mechanism during their spasms corroborates the differences found in the eye-movement records. These conclusions can be made only tentatively until they are verified by like measurements on apparatus which obtains both vertical and horizontal movements of both eyes and which also makes

possible a more accurate identification of points on the record and particular disturbances in the speech mechanism.

Only four of our most severe stutterers showed any indication of binocular incoördination during spasms. Three of these showed a very interesting vertical twitching of one eye while the other eye remained in a relatively rigid fixation. The other stutterer showed incoördinated independent horizontal movements of the two eyes; a kind of temporary strabismus. No attempt will be made to explain these phenomena until our research in this line has reached a more advanced stage.[3]

A very interesting comparison may be made between the nature of the deviations of the stutterers from normal speakers and the nature of the deviations found by Diefendorf and Dodge in the different insanities (**2**). Speaking of the eye-movement records of maniacal patients they state (p. 463) that "it is again congruent with the disease-picture that the eye-movements, which we have found to be rapid, are secondary automatic acts, not those that require conscious direction and control, such as have hitherto been measured; and it certainly corresponds closely with our general knowledge of the diffusion of the sensory impulses and the interaction of the higher and lower centers, that these secondary automatic movements should reach their extreme velocity when the interaction of the higher nervous centers is lessened." These authors posit that the slow movements found in depressed and epileptic patients must be due to an involvement of the lower centers (centers for the simple reflexes and automatic acts) as well as the higher centers. The grossly demented patients showed marked inaccuracies of fixation and the patients with advanced dementia paralytica showed 'curious inconsequential fixations, breaking the normal eye-movements at irregular points.'

[3] As this paper goes to press the senior author has just obtained some records with the new Iowa camera which give conclusive evidence of binocular incoördination in five of our most severe stutterers. Three of the subjects showed incoördination only in vertical movements, one eye remaining fixed during a spasm while the other moved markedly up or down or one eye moving up during a spasm with a coincident downward movement of the other eye. The horizontal incoördination was characterized by either a complete independence of movement of the two eyes or a coincident extreme convergent strabismus.

The eye-movements of stutterers during spasms of the speech mechanism are interestingly and remarkably well described by Diefendorf and Dodge in their discussion of the eye-movements of insane patients who have marked disturbances of higher center control with consequent increase in lower center activity. The rapid eye-movements of the stutterers during clonic spasms, together with the marked vertical twitches is a convincing demonstration of the breaking through of lower center processes. The increased number of fixations per line with no significant increase in average duration of fixation in spite of the several extremely long fixations of the stutterers is considered additional evidence that the nervous disturbance in stuttering falls into the class of disturbances caused by a decrease in cortical control over lower center processes. This finding bears out the work of Travis (**11**) on the decrease of reflex time during stuttering, and of Herren (**12**) on the decrease in extent of tremors during stuttering.

General Summary and Conclusions

By means of photographic records the eye-movements of 15 stutterers during oral reading were compared with like movements of normal speakers. Comparisons were also made between the differences in measurements of eye-movements in oral and in silent reading of stutterers and normal speakers. Qualitative analyses of the movements of the eyes during stuttering spasms were compared to like analyses on the insane by Diefendorf and Dodge. The results obtained indicate the following conclusions:

1. Stutterers in oral reading present more regressions per line, more fixation pauses per line, and about the same average duration of fixation as do normal speakers.

2. Stutterers show greater differences between silent and oral reading in all quantitative measurements of eye-movements than do normal speakers.

3. The finding of disturbances in eye-movements coincident with and analogous to disturbances in the speech mechanism suggests a general disequilibration of the nervous system as a whole during stuttering.

4. Since the variations in the movements of the eyes during spasms of the speech mechanism seem to be somewhat analogous to the variations in types of spasms, photographic records of eye-movements during speech spasms may be a means of differentiation types of spasms and possible types of stutterers.

5. The automatic or reflex type of eye-movement found during stuttering suggests a relative decrease in the control of cortical over sub-cortical centers.

6. Interpretatively this study may be considered as supporting the concept emphasized by Travis (**11**) that the act of stuttering may be envisaged as a temporary recurrent reduction in the dynastic control of the superjacent levels over their substructures causing general disequilibration of the orientation of potentials within the nervous system necessary for efficient speech.

REFERENCES

1. DEARBORN, W. F., The psychology of reading, *Arch. of Phil.*, and *Scien. Method*, 1906, **4**, Columbia Contributions 14, No. 1.
2. DIEFENDORF, A. R., AND DODGE, RAYMOND, An experimental study of the ocular reactions of the moone from photographic records, *Brain*, 1909, **31**, 452–489.
3. DODGE, R., An experimental study in visual fixation, *Psychol. Rev., Monog. Suppl.* 1907, **8**, 95.
4. GRAY, C. F., Types of reading ability as exhibited through tests and laboratory experiments, *Elem. Educ. Monog.* Univ. of Chicago, no. 1, Aug. 1917.
5. GRAY, W. S., *Summary of Investigations Relating to Reading*, Univ. of Chicago, Monographs on Educ., no. 28, June 1925.
6. HUEY, E. B., *The Psychology and Pedagogy of Reading*, Macmillan, N. Y., 1908; also, *Amer. J. of Psychol.*, April 1901, **12**, 292–312.
7. MILES, W. R., AND SHEN, E., Photographic recording of eye movements in the reading of Chinese in vertical and horizontal axes, J. EXP. PSYCHOL., 1928, **8**, 344–362.
8. MURRAY, ELWOOD, Disintegration of breathing and eye movements in stutterers during silent reading and reasoning, Thesis, Univ. of Iowa, Feb. 1931; to be published as an Iowa Study of the *Psychol. Monog.* in 1931.
9. SCHMIDT, W. S., An experimental study in the psychology of reading, *Suppl. Educ. Monog.* Univ. of Chicago, April 1917, **1**, no. 2.
10. TINKER, M. S., Eye movements, duration, pause duration, and reading time, *Psychol. Rev.*, 1928, **35**, 385–397.
11. TRAVIS, LEE E., *Speech Pathology*, N. Y., D. Appleton Co., 1931.
12. HERREN, R. Y., The effect of stuttering and alcohol upon certain tremor frequencies, J. EXP. PSYCHOL., in press.

(Manuscript received June 15, 1931)

HABITUATION, RETENTION, AND PERSEVERATION CHARACTERISTICS OF DIRECT WAKING SUGGESTION *

BY EVERETT F. PATTEN, ST. CLAIR A. SWITZER
Miami University

AND

CLARK L. HULL
Yale University

I

In 1931 Krueger published a study (5) which showed that the rate of lid closure in response to hypnotic suggestions increased markedly and progressively in speed with repetition, and that the practice effects thus acquired deteriorated perceptibly from a 90-day period of no-practice. In general, Krueger found hypnotic lid closure conforming to the classical characteristics of habituation. The preceding year, Hull and Huse (4) published a graph showing typical practice effects in responding to postural suggestions. In 1931, Williams (7) published several such graphs. Both these latter studies were primarily directed to the solution of quite different problems which, unfortunately, required that the subjects receive their practice partly in the trance and partly in the waking state. Moreover, no tests were given to determine retention after an interval of no-practice. It was accordingly felt desirable to secure clear and unambiguous evidence on waking postural suggestion parallel to that of Krueger on hypnotic lid closure.

Sixteen college students, seven young men and nine young women, who were sufficiently suggestible to fall completely forward in the posture test when in the waking condition, served as subjects. None of these people had ever been hypnotized, except subject Ro, who had been hypnotized once,

* From the Psychological Laboratory, Miami University. The experimental work was performed in about equal parts by the two junior authors. For the design, formulation, and interpretation the senior author is mainly responsible.

TABLE I

THE NUMBER OF SECONDS REQUIRED ON SUCCESSIVE DAYS TO PRODUCE COMPLETE FALLING FORWARD BY WAKING SUGGESTION

Trial No.	Subjects																Mean Time
	Ba	Ch	Fe	Fl	Os	Mo	We	Bo	Co	Da	Fi	Go	Ki	Mc	Ro	St	
1	22.6	40.8	65.8	24.2	41.4	59.2	28.6	37.1	35.3	55.4	16.4	30.0	18.6	19.8	32.0	24.4	34.47
2	23.4	23.6	57.8	13.8	54.6	46.8	27.8	24.6	14.2	70.8	10.5	15.0	10.2	14.6	20.5	8.4	27.28
3	21.2	25.2	28.8	9.2	65.8	24.8	56.8	21.8	11.9	22.5	8.1	17.0	18.8	13.5	19.1	9.5	23.37
4	18.6	19.4	24.8	3.6	68.0	25.2	39.6	17.0	10.2	24.8	8.6	12.8	24.0	15.2	21.8	6.2	21.23
5	17.0	17.4	21.4	14.0	68.0	27.6	78.8	13.2	9.8	22.6	6.4	12.5	9.1	15.5	19.8	11.0	22.75
6	20.4	13.4	18.2	8.6	37.0	25.2	52.8	17.0	7.5	20.5	5.1	6.8	12.1	12.1	30.4	7.4	18.40
7	14.2	17.4	36.2	7.2	38.0	24.0	63.8	16.4	6.9	15.6	6.7	10.0	24.8	15.4	25.2	7.4	20.57
8	14.0	11.8	24.8	7.2	30.0	26.4	43.4	14.6	7.4	11.6	6.7	8.1	18.8	11.6	22.7	8.9	16.75
9	12.0	13.4	23.8	7.8	32.6	22.4	29.8	15.8	5.7	8.5	5.9	7.8	21.6	7.5	19.4	7.2	15.07
10	17.2	13.4	19.0	8.2	22.8	23.4	32.2	12.3	7.8	11.7	5.4	8.4	15.0	10.6	18.1	7.6	14.56
11	18.2	16.6	23.2	8.2	30.2	14.2	28.4	14.4	5.2	6.5	5.1	11.0	12.5	9.2	16.0	7.9	14.17
12	15.4	14.2	15.8	7.2	37.2	19.6	28.2	10.7	6.8	11.0	5.6	8.8	12.9	8.3	14.3	9.2	14.07
13	17.2	10.0	17.0	8.4	29.6	17.6	13.4	14.0	7.0	12.3	5.3	9.5	13.2	8.5	14.0	14.3	13.20
14	16.0	11.2	16.4	8.4	22.4	16.8	19.2	12.3	5.2	6.4	5.4	9.7	12.4	8.1	10.0	9.7	11.85
15	18.2	9.8	19.2	6.8	19.6	15.0	16.4	9.6	5.4	4.3	5.5	7.8	9.0	8.8	6.9	7.2	10.59
16	12.8	9.6	12.4	10.6	14.2	14.2	15.0	9.4	7.0	3.5	4.6	10.1	9.0	8.0	14.2	8.2	10.17
17	15.0	8.2	17.2	7.0	17.2	17.0	12.0	7.5	5.1	3.1	6.3	9.6	7.7	7.0	13.1	4.8	9.86
18	18.8	7.8	11.2	8.0	15.6	15.2	10.0	9.6	4.9	6.0	4.9	12.5	11.5	8.4	14.4	6.2	10.31
19	13.0	7.4	11.2	6.2	14.8	15.0	11.2	10.4	5.8	3.3	3.5	10.5	7.4	11.3	12.5	6.0	9.34
20	14.4	8.8	10.6	7.8	18.4	11.8	9.4	10.6	5.3	3.4	3.8	7.4	8.7	8.5	10.6	8.6	9.25
Seventy-day interval of no practice																	
21	12.1	6.3	20.8	4.5	24.1	17.1	31.4	12.6	7.7	7.4	5.0	14.4	19.2	13.8	16.9	14.4	14.23
22	11.8	6.4	21.8	9.1	24.9	10.1	29.2	8.6	7.0	6.8	3.4	10.9	11.9	11.3	14.5	7.9	12.22
23	17.3	5.7	21.1	5.9	20.6	13.0	26.4	10.3	6.2	11.3	4.1	9.8	9.5	13.3	14.2	6.5	12.20
24	19.9	7.1	27.2	6.9	19.1	10.1	15.2	10.3	8.0	10.6	4.9	10.6	12.3	12.1	12.3	6.0	12.03
25	13.5	6.7	23.4	5.5	19.5	10.9	26.1	8.7	6.4	6.6	4.2	8.7	18.5	9.6	10.0	7.2	11.59
26	13.9	7.2	16.1	3.2	18.3	12.3	17.2	8.6	5.2	6.6	3.9	10.2	15.2	8.7	11.4	4.7	10.16

two years previously. The Hull postural suggestion technique (3, 153 ff.) was administered to each subject three times at one-minute intervals on each of 20 successive days. Seventy days after the conclusion of the above, practice was resumed in the same manner for six successive days. There were thus obtained a total of 1,248 suggestion measurements in the form of the time required for spoken suggestions to induce the standing subject to fall far enough forward to lose his balance. The postural movements were recorded graphically, and the times were transcribed from the smoked-paper records. The greatest care was exercised throughout the experiment to have the suggestions constant as to wording, tempo, and emphasis. The same experimenter administered all the suggestions to a given subject.

The time required for the first response to suggestion on each of the 26 days by each of the 16 subjects is shown in Table I. In the last columns at the right are given the means for all 16 subjects taken together on each of the 26 days of

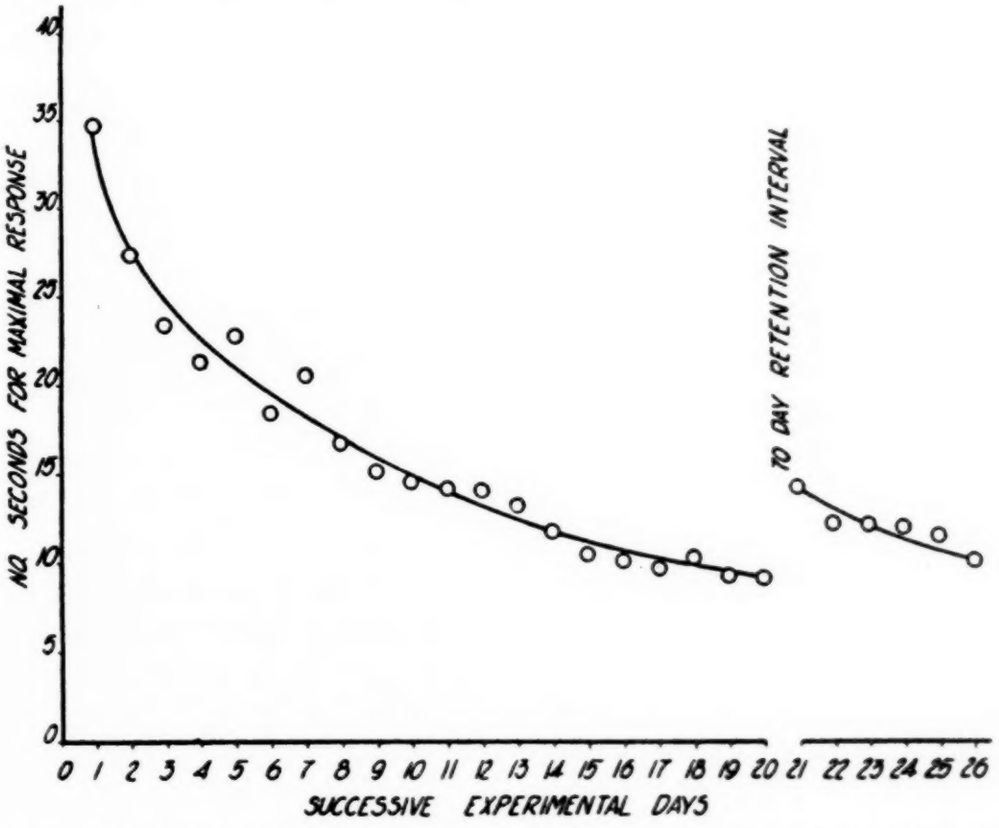


FIG. 1. Composite graph from 16 subjects showing practice, retention, and recovery phenomena of waking postural suggestion.

practice. From these means there have been plotted the composite practice curves shown in Fig. 1. An inspection of this figure shows at once that the behavior of waking postural suggestion conforms in all measured respects to the laws of habituation:

1. Successive repetitions radically reduce the time of performance.

2. The rate of reduction is more rapid at first, and slower later, generating a curve of negative acceleration.

3. Seventy days of no-practice results in a marked loss of the facility produced by previous repetitions.

4. The resumption of practice is followed by a recovery of the loss evident after the period of no-practice. There is a little indication that the rate of recovery may be more rapid than was its original acquisition.

II

There is little doubt but that states of heightened suggestibility exist. Both Hull and Huse (4) and Williams (7) have shown that postural suggestions take place considerably more quickly in the trance state than in the waking condition, thus indicating that the trance, as has so long been supposed, is indeed a state of heightened suggestibility. Hull and Huse also found that postural suggestions took place perceptibly more promptly 15 minutes after waking from the trance than preceding hypnosis. This indicates not only that hypnosis is a state of heightened suggestibility, but that this heightening perseverates for some time after the trance has been formally terminated. Finally, the principle of perseveration of hypnotic hypersuggestibility has been confirmed by Krueger's observations that lid closure in response to hypnotic suggestion immediately after waking from a trance takes place in approximately half the normal time, and that an appreciable amount of hypersuggestibility seems to be present a full hour after dehypnotization.

Having established the fact of hypersuggestibility, we are immediately faced with the question of its nature. Is hyper-

suggestibility a unique thing or is it merely a special case of some more familiar process? The results of certain recent investigations point in the latter direction. Hull and Huse noticed, for example, that the response to postural suggestions, in both the trance and waking states, was distinctly more prompt immediately after having responded to a suggestion of the same kind. They take the tentative position that the response to a suggestion (*any* suggestion) may be the essential mechanism which brings about states of hypersuggestibility. Thus, responding to a suggestion to fall forward leaves the organism with a heightened susceptibility to a second postural suggestion, actually reducing the period required for maximal response by about a fourth; having responded to a suggestion to go to sleep by closure of the lids and other less observable behavior, the organism is left with a heightened susceptibility to respond not only to suggestions of sleep by lid closure, but to postural suggestion as well, actually reducing the suggestion time by about a half. According to this hypothesis, then, the hypersuggestibility of the hypnotic trance is no different in kind from that generated by responding to any other suggestion.

But further complications at once arise when the attempt is made rigorously to apply the above hypothesis to the detailed experimental phenomena of hypnosis and suggestibility. One major difficulty of this kind lies in the fact that ordinarily the response to one suggestion not only facilitates the response to an immediately succeeding one, but an appreciable amount of such facilitation persists 24 hours or longer. But this latter facilitation, by all ordinary criteria, is a practice or habituation effect. It is clear, then, that so long as the subject is in a state of practice, as shown by a gain in facility of response to suggestion after 24 hours, any immediate increase in facility from such a cause must consist, to a certain extent at least, of practice effects.

But if an appreciable amount of genuine practice effects persists for 24 hours, there is excellent reason to believe that there must have been a very much larger amount of such effect immediately after the practice. We know that some forms of

habituation weaken very rapidly with the passage of time, particularly at first. In the specific field of hypnosis, Krueger has published evidence (6) which suggests that hypnotic hypersuggestibility follows this tendency after the subject wakes from the trance. Ebbinghaus, in his classical study of forgetting (2), found that the amount retained after 24 hours was only about a fifth of the learning manifest immediately at the close of the practice. If hypnotic practice effects are as unstable as Ebbinghaus found just barely learned nonsense syllables to be, we might assume that what appears to be hypersuggestibility immediately after waking from a trance or responding to some other form of suggestion, contains something like five times as much of actual practice effects as survive until the next day. This naturally raises the question of whether there is in what has been called hypersuggestibility any element at all beyond ordinary practice effects. In short, there emerges the hypothesis that induced hypersuggestibility in general, and hypersuggestibility characteristic of the hypnotic trance in particular, are in reality but special cases of ordinary practice effects. In thus reducing two supposedly distinct phenomena to a single principle, the present hypothesis has at least the virtue of conforming to the principle of parsimony.

But how shall we test this hypothesis? One obvious approach to the experimental solution of the problem is to test for hypersuggestibility some subjects who have reached the limit of practice, *i.e.*, a point beyond which practice no longer produces any permanent improvement in the function in question. The hypothesis demands that the phenomenon of hypersuggestibility in the particular function should be completely lacking in such individuals. And if the amount of hypersuggestibility should be tested frequently throughout the practice period, this would be found to lessen progressively as the practice curve (Fig. 1) approaches the horizontal. Accordingly, each of 16 individuals was put through the postural suggestion procedure three times at one-minute intervals on each of 20 successive days, and again after 70 days of no-practice, for six successive days, as pointed out in the

previous section. It was hoped that these persons would have reached the limit of practice some time before the end of this rather protracted period of exercise The hypothesis demands that if separate composite graphs be plotted of the time required for the first, second, and third falls throughout the 20 days of practice, these curves should be rather far apart on the first days, they should converge more and more as practice goes on, and, at the limit of practice, they should actually coincide. But if the hypothesis is incorrect and there is involved in the immediate facilitation of suggestibility ordinarily following a positive response to suggestion a genuine hypersuggestibility quite distinct from practice effects, the lines of the graph should either not converge at all or they should converge at a slower rate than the rate of diminution in the permanent results of practice. According to the latter alternative, at the limit of practice, where the practice curve becomes quite horizontal, the three lines should still be separated by an appreciable amount. Except for distortions due to sampling, the extent of this separation should represent the amount, if any, of genuine hypersuggestibility as distinguished from immediate but true practice effects.

The mean values in terms of suggestion time of the 16 subjects on each of the 26 days, and for each of the three successive trials per day, are shown in Table II. These values are presented graphically in Fig. 2. A glance at this graph shows at once that the curves of the three successive falls definitely converge. In so far the hypothesis is confirmed.

Unfortunately, the expectation that the limit of practice in this function would be reached several days before the 20th day was not realized. Consequently the criterion of complete coincidence of the curves can not be applied. A secondary criterion remains, however. This is that the rate of convergence should be slower than the diminution in rate of learning. This means that the per cent of the immediate gain which persists for a period of 24 hours, which is obviously considerable during the early stages of practice, should decrease at a fairly rapid rate until, at the limit of practice, it should obviously become zero. Appropriate computations

TABLE II

MEAN SUGGESTION TIME FROM 16 SUBJECTS AT EACH OF THE THREE SUCCESSIVE FALLS FOR EACH OF THE 26 DAYS OF THE EXPERIMENT

Day No.	Mean Number of Seconds of Suggestion Required to Produce the "Fall"		
	First Fall	Second Fall	Third Fall
1	34.47	26.20	22.53
2	27.28	21.36	17.77
3	23.37	17.86	17.16
4	21.23	19.45	16.66
5	22.75	18.16	16.52
6	18.40	14.81	14.43
7	20.57	14.80	14.81
8	16.75	13.24	12.49
9	15.07	14.64	12.74
10	14.56	13.40	11.61
11	14.17	11.07	11.30
12	14.07	10.87	10.10
13	13.20	11.85	12.31
14	11.85	11.83	9.83
15	10.59	9.75	9.69
16	10.17	9.03	8.84
17	9.86	9.06	7.89
18	10.31	8.66	8.93
19	9.34	8.67	8.51
20	9.25	8.06	8.10
Seventy-day Interval			
21	14.23	12.17	11.94
22	12.22	10.95	11.26
23	12.20	9.82	9.45
24	12.03	10.43	9.31
25	11.59	10.50	8.83
26	10.16	8.17	8.83

from the data of Table II show that, on the average, there is discernible some indication of such a tendency. The percentages of immediate facilitation surviving for 24 hours in the respective fourths of the practice period are as follow:

First Period	Second Period	Third Period	Fourth Period	Recovery Period
36%	36%	31.5%	21%	35%

The percentage of immediate facilitation surviving 24 hours during the last fourth of the practice is slightly over half that of the first fourth. However, the recovery period, which corresponds roughly to the last fourth of the original practice

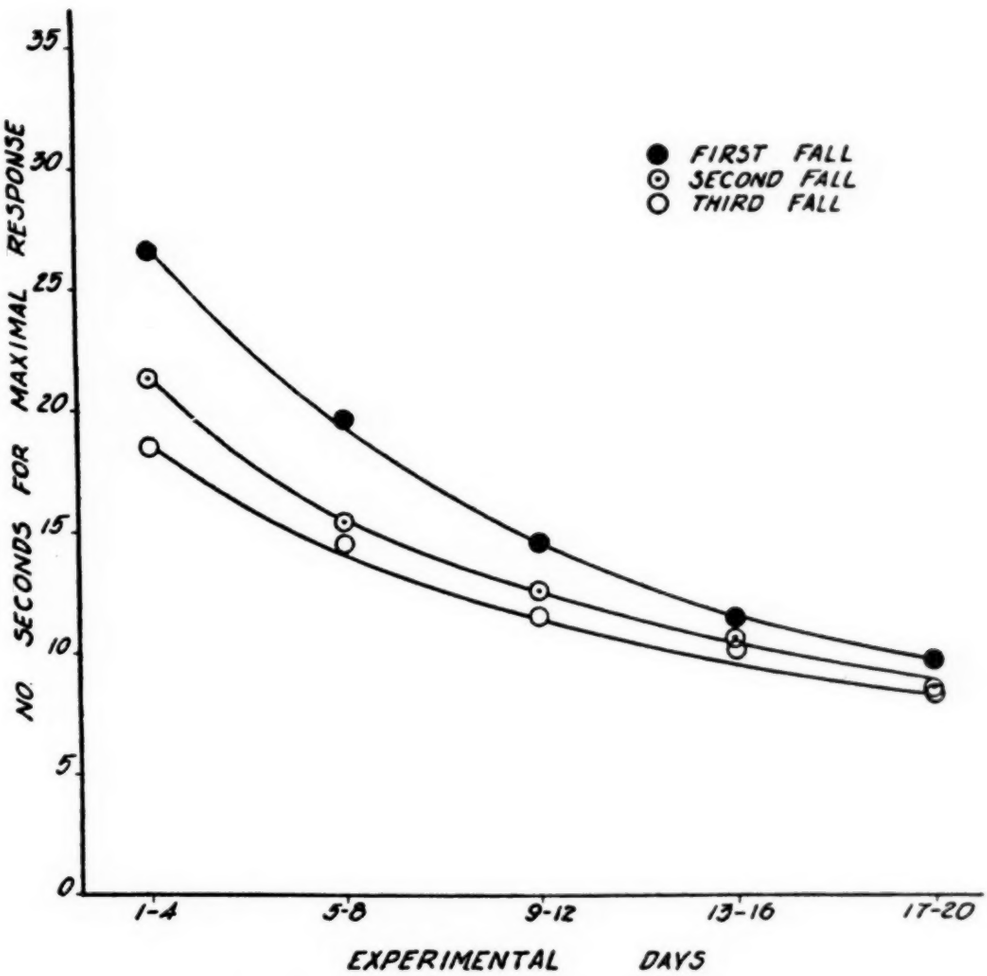


FIG. 2. Graph showing the diminution in the immediate facilitation of suggestibility as the amount of permanent facilitation decreases with practice.

series, again manifests practically the original per cent of survival. It is difficult to say how much weight in this matter should be given to the results from the recovery period, because its exact status is not yet fully known in ordinary learning phenomena, to say nothing of our general ignorance of the function among suggestion phenomena. If it be disregarded, the ordinary practice series seems to show a fairly marked and consistent tendency pointing in the direction of a residual of genuine hypersuggestibility distinct from practice effects. It is clear, however, from the marked convergence of the curves in Fig. 2, that in case such a hypersuggestibility actually exists it must be very small in amount. It accordingly seems likely that a considerable part, at least, of the

hypersuggestibility observed in ordinary hetero-suggestion experiments is in reality a more or less transitory practice effect.

Summary and Conclusions

The latency of maximal response to postural suggestion in the waking state decreases progressively for a period of 20 days. The practice curve manifests a marked negative acceleration. A 70-day period of no-practice produces a marked increase in the latency which is largely recovered from during a subsequent 6-day resumption of practice. It is believed that these phenomena, considered in conjunction with the experimental conditions under which they were produced, indicate that this form of waking suggestion is primarily a habit phenomenon. Furthermore, the fact that the present results parallel so closely those of Krueger on hypnotic lid closure strongly supports the general *a priori* probability that the two processes are essentially of the same nature.

About a third of the amount of increase in the speed of maximal response to postural suggestion in the waking state is found to persist for 24 hours. Bean's summary (1, 40 ff.) shows that it is not unusual for two thirds of immediate practice effects to be lost during such a length of time. This suggests that what appears superficially as hypersuggestibility may be nothing but immediate, though somewhat transitory, practice effects. This hypothesis is confirmed by the fact that the immediate gains from practice decrease progressively as the limit of practice in the function is approached. It is clear that if there is any element of hypersuggestibility in the immediate practice effects under the conditions of the present experiment, it must be relatively slight. If no such element whatever exists, the hypothesis would seem to demand that the ratio of the permanent facilitation should be constant, whereas the ratio should decrease, and become zero at the limit of practice, if there is an appreciable residue of genuine hypersuggestibility as distinct from practice effects. There is some indication that the ratio tends toward zero as practice goes on.

By way of caution it should be pointed out that if a very

small residue should be found at the limit of practice, this may be significant because it should correspond to a very much larger absolute amount during the early stages of practice, in which condition subjects are often experimented on. It should also be emphasized that only one of two important cases of hypersuggestibility has been here investigated. The case here considered concerns the influence of responding to a suggestion upon the facility of responding to another of the same kind soon after. The second case is that encountered in hypnosis where the response to one suggestion (sleep, lid-closure) is supposed to facilitate response to a *different* suggestion, such as postural movement. If the first case be considered in the category of practice effects, the second must obviously fall under the head of transfer of training. If the above hypothesis is sound, there should be a transfer (doubtless with diminution) from one hetero-suggestion to any other. In particular, hypnotic lid closure ought to take place more readily after a positive response to a postural suggestion, just as postural suggestion takes place more readily after positive response to sleep suggestions by lid closure. Numerous other experimental approaches of the same general nature are possible. It is to be hoped that this critical matter may be cleared up by experiment at an early date.

References

1. Bean, C. H., The curve of forgetting, *Arch. of Psychol.*, No. 21, March, 1912.
2. Ebbinghaus, H., Memory, a contribution to experimental psychology, Translated by Ruger, H. A. and Bussenius, C. E., Teachers College, Columbia University, 1913.
3. Hull, C. L., Quantitative methods of investigating waking suggestion, *J. Abn. and Soc. Psychol.*, XXIV, 1929, 153–169.
4. Hull, C. L. and Huse, B., Comparative suggestibility in the trance and waking states, *Amer. J. Psychol.*, XLII, 1930, 279–286.
5. Krueger, R. G., The influence of repetition and disuse upon rate of hypnotization, J. Exper. Psychol., XIV, 1931, 260–269.
6. Krueger, R. G., Variation in hypersuggestibility preceding, during, and following the hypnotic trance, *J. Abn. and Soc. Psychol.*, XXVI, 1931, 131–140.
7. Williams, G. W., Suggestibility in the normal and hypnotic states, *Arch. of Psychol.*, No. 122, 1930.

(Manuscript received April 16, 1932)

IS THE JUDGMENT OF THE GROUP BETTER THAN THAT OF THE AVERAGE MEMBER OF THE GROUP?

BY J. RIDLEY STROOP[1]
George Peabody College

Goodwin B. Watson (1928), after stating that the experimental evidence in this field is limited, says: "Miss Gordon has shown that the average of the individual judgments in such matters as esthetics or in the judgment of a lifted weight is equal or superior to the judgment of the best individual within the group. . . . This is encouraging evidence on the quality of the group product." His reference is to Miss Gordon's studies (1923 and 1924) on ranking pictures for their beauty and on ranking weights. These statements have been given further publicity by their incorporation of Watson's article in Wheeler's Readings in Psychology (1930).

Since Gordon's studies have formed the basis for such far-reaching statements, and the results of one of them (1924) have been used by Kelley (1925) as evidence of the reliability of the Spearman-Brown prediction formula, it seems fitting that they should receive further attention experimentally. Though her two studies are very different in character, the apparently convincing results are due largely to the operation of the same factor: the mathematical effect of her method of grouping. For this reason attention in this paper will be given only to the second study.

Since Gordon's procedure is fundamental to the present experiments, it is described here. She used a series of 10 weights ranging from 16 grams to 17.6 grams, with the difference between the successive weights (1/90 of the lightest weight) equal throughout. Each of 200 subjects placed the

[1] The writer wishes to express publicly his sincere appreciation to Dr. Joseph Peterson, under whose direction this study was made, for his kind assistance at all times and especially for his criticism of this manuscript, and to Dr. Lyle H. Lanier for suggestions and coöperation in obtaining subjects for Parts I and II.

weights in order from the heaviest to the lightest, being allowed sufficient time to make any rearrangements desired. The final order resulting from each individual's ranking was correlated with the true order of the weights by the rank-difference method. The average of the 200 coefficients of correlation, which was taken to represent the value of the judgment of the average member of the group, was .41. The judges were then combined into mutually exclusive groups of 5, and a composite-judgment order of weights for each group was made by averaging the positions assigned each weight by the several members of the group. These composite judgments were ranked and this ranking was then correlated, for each group, with the true order of the weights. The average of these coefficients was taken to represent the value of the group judgment by a group of 5. By the same procedure values of group judgments by groups of 10, 20, and 50 were calculated. The validity found in the judgment by groups of 5, 10, 20, and 50 was .68, .79, .86, and .94, respectively, as compared with .41, the mean validity of the judgments by single individuals.

Experimental

The present study is composed of five parts. (1) An experiment was made using 100 subjects following Gordon's technique exactly. This served as a check on her results and also as a basis for comparing general experimental conditions. (2) Judgments[2] by the same individual were combined according to Gordon's method of grouping. The number of individuals used corresponded exactly to the number of groups averaged by Gordon. The number of rankings made by each individual corresponded exactly to the number of subjects in each of Gordon's groups. The results of each individual's rankings were averaged as were the results of Gordon's groups. This required the same number of mathematical operations as she employed thus equating the mathematical effects of averaging. All other conditions were held exactly comparable with those described by Gordon. (3) The average validity of the judgments was increased by doubling

[2] The term 'judgment' as used by Gordon and also by the writer is synonymous with ranking.

the range of the series of weights, making the increment from the lightest to the heaviest 1/5 of the lightest weight. Each weight differed from its neighbor by 1/45 of the lightest weight. The lightest weight was unchanged (16 grams). The effects of grouping judgments or rankings both by the same subject and by different subjects are shown in comparison. (4) Arrangements of weights were determined by chance and the orders resulting were correlated with the true order of the weights. This eliminates the judgment factor and of course increases the chance factors. (5) The actual increase in validity obtained by grouping is compared with that predicted by the Spearman-Brown formula. These will be discussed separately.

1. *A Check on Gordon's Results*

As Gordon's original data were not published, a short study was made as a check on her results. One hundred students in attendance at George Peabody College during the summer term of 1930 served as subjects. The materials used and the methods employed were in exact agreement with the description above; the only variation being the number of subjects. The results compare favorably with those of Gordon's experiment as shown in Table I.

TABLE I

A Comparison of the Means of the Coefficients from This Part of the Present Study with Gordon's

Size of Group	1	5	10	20	50
Gordon's results	.41	.68	.79	.86	.94
Our results	.38	.68	.85	.92	.96
Difference	.03	.00	.06	.06	.02

2. *The Individual Substituted for the Group*

In this section the individual is substituted for the group: *i.e.* instead of using 200 subjects and combining the judgments or rankings into 40 groups of 5 each, 20 groups of 10 each, 10 groups of 20 each, and 4 groups of 50 each, as was done by Gordon; 40 subjects ranked the weights 5 times each, 20 subjects 10 times each, 10 subjects 20 times each, and 4

subjects 50 times each. The 40 subjects were selected at random from the general psychology class which was composed of 54 members. The 20 subjects were selected at random from the 40, the 10 from the 20, and the 4 from the 10. Five rankings were made at one sitting and never were more than two sittings required per day: one in the forenoon and one in the afternoon. Three sets of weights were employed and shifted so that no subject used the same set twice in succession during the same sitting.

The subjects were asked to arrange the weights in order from the heaviest to the lightest, each time depending wholly on the information obtained by lifting them. To encourage the subjects to follow these directions closely and not look for other clues their attention was called to the fact that there was nothing about the weights that would give a clue to their true positions. Also comment was made upon the small difference between the weights, and each subject was allowed to see his record for a few rankings near the beginning to furnish evidence that only occasionally was a weight being put in its proper place and that many were far removed from the true order. This served to deter an effort to 'spot' weights or return them to the same positions. However, through subjects' report one irregularity was found (no other). One subject placed one weight one time by the keyed symbol on its bottom. This irregularity was considered negligible and ignored. Each subject spent from five to fifteen minutes upon each ranking. Many subjects did not look at the weights but gazed into space while judging, in an effort to avoid visual distraction.

The question of the improvement due to learning made by those who arranged the weights a large number of times was measured by comparing the mean of the coefficients indicating the validity of the first 15 rankings and that representing the validity of the last 15 rankings for each of the four subjects who ranked the weights 50 times. Two subjects showed gains of .043 and .030 respectively, while the other two subjects showed losses of .026 and .094 respectively. The repetition of the individual judgment had no appreciable effect upon the value of the judgment.

The data secured when the individual was substituted for the group are given in Table II. Column 1 is composed of the numbers that represent the subjects. Column 2 was obtained

TABLE II

COEFFICIENTS OF CORRELATION SHOWING THE IMPROVEMENT IN THE VALIDITY OF COMPOSITE JUDGMENTS OVER INDIVIDUAL JUDGMENTS BY THE SAME SUBJECTS

Number of Subject	Means of r's of 1st 5 Single Judgments	r's of Composite Judgments for Each Subject. The Number of Rankings Grouped			
		5	10	20	50
1	.41	.94	.85	.84	.95
2	.21	.48	.65	.87	.96
3	.26	.75	.82	.94	.98
4	.47	.65	.85	.94	.98
5	.57	.90	.85	.90	
6	.47	.71	.74	.76	
7	.67	.87	.94	.98	
8	.07	.13	.68	.84	
9	.29	.56	.55	.70	
10	.40	.72	.94	.89	
11	.52	.94	.94		
12	.52	.98	.96		
13	.63	.83	.93		
14	.33	.61	.84		
15	.44	.78	.95		
16	.56	.71	.69		
17	.57	.82	.87		
18	.39	.66	.52		
19	.27	.46	.84		
20	.55	.84	.91		
21	.18	.39			
22	.72	.93			
23	.24	.54			
24	.34	.88			
25	.22	.56			
26	.76	.91			
27	.70	.89			
28	.23	.50			
29	.60	.90			
30	.62	.71			
31	.22	.62			
32	.28	.42			
33	.54	.97			
34	.54	.84			
35	.34	.74			
36	.29	.61			
37	.52	.76			
38	.41	.84			
39	.39	.64			
40	.46	.81			
Totals.....	17.20	28.80	16.32	8.66	3.87
Means.....	.43	.72	.82	.87	.97
No. Sep. J.....	200	200	200	200	200

by correlating each of 5 judgments for the several subjects with the true order of the weights and calculating the mean for each group of 5. Column 3 shows the coefficients of correlation representing the values of composite judgments secured by combining, according to Gordon's procedure described above, 5 separate judgments for each of 40 subjects. Column 4 shows the coefficients which represent the values of composite judgments for 20 subjects which resulted from combining 10 separate judgments by the respective individuals. Column 5 gives the values of composite judgments for 10 subjects based upon 20 judgments by each subject, and the last column (**6**) shows it for 4 subjects when 50 judgments by each subject are combined.

In the final results, the second row from the bottom of the table is composed of means of the data found in the respective columns above. These means are taken to represent the validities of individual judgments as obtained by grouping different numbers of rankings. The individual judgment based upon one ranking of the weights has a validity of .43. The individual judgment based upon groupings of 5 rankings has a validity of .72; upon 10 rankings, .82; upon 20 rankings, .87; and upon 50 rankings, .97. The bottom row of the table indicates that the same total number (200) of rankings (though in different groupings for averaging) were used throughout. This was done so that the number of mathematical operations employed would correspond to the number used by Gordon.

Table III shows, in frequency distributions, a comparison between Gordon's results and those obtained when the groups were formed of rankings or judgments by the same individual. Where the judgments grouped were made by the same individual practically the same increase in validity resulted from grouping as that obtained when the judgments were made by different individuals. As this substitution of judgments by the same subject for judgments by different subjects is the only point of difference between the two studies, and as it causes no appreciable difference in the results, it seems reasonable to conclude that using judgments by different subjects is not a vital factor.

TABLE III

A Comparison of the Effects of Combining Rankings Made by Different Subjects with the Effects of Combining Corresponding Numbers of Rankings Made by the Same Subjects

Frequencies of Coefficients Showing the Relation between Rankings by Groups of Different Sizes and the True Order of the Weights. (Gordon's Data)						Frequencies of Coefficients Showing the Correlation between Rankings by the Same Subject Based upon Different Numbers of Single Rankings and the True Order of the Weights. (Our Data)				
Size of Groups					Coefficients	Number of Rankings				
50	20	10	5	1		1	5	10	20	50
4	5	3	1	5	90 to 99	5	8	7	4	4
	3	10	12	12	80 to 89	14	9	7	4	
	1	5	13	20	70 to 79	24	8	1	2	
	1	1	5	20	60 to 69	23	6	3		
		0	2	35	50 to 59	21	4	2		
		0	3	29	40 to 49	25	3			
		0	2	15	30 to 39	25	1			
		0	1	21	20 to 29	17	0			
		1	0	11	10 to 19	16	1			
			1	8	0 to 9	14				
				8	−10 to − 1	7				
				6	−20 to −11	5				
				2	−30 to −21	3				
				3	−40 to −31	0				
				1	−50 to −41	0				
				3	−60 to −51	0				
				0	−70 to −61	0				
				0	−80 to −71	0				
				1	−90 to −81	0				
.94	.86	.79	.68	.41	Means	.43*	.72	.82	.87	.97

* The means on the right side of this table were taken from Table II.

3. *The Effects of Grouping Judgments Made on More Easily Discriminable Weights*

The subjects for this part of the study were 100 students enrolled in the two years of junior college work at David Lipscomb College during the fall quarter 1930. The series of weights employed ranged in weight from 16 grams to 19.2 grams with a uniform increment throughout of 1/45 of the lightest weight. Five sets of weights were used and these were shifted for those who ranked them more than once, to destroy the influence of any possible 'ear-mark.' At least one hour lapsed between any two arrangements by the same subject.

First, each subject arranged the weights from the lightest to the heaviest and the results were correlated with the true order of the weights. This process yielded 100 coefficients the mean of which was .72. The 100 rankings were then combined into 50 groups of 2, 20 groups of 5, and 10 groups of 10; and correlations were obtained between the several orders resulting from these groupings and the true order by the procedure outlined in the introduction. The means of the coefficients which were taken to represent the validity of the judgments by groups of 2, 5, and 10 were .83, .92, and .96, respectively.

Next, 50 subjects, selected at random from the 100, ranked the weights 2 times each; 20 subjects, selected at random from the 50, ranked them 5 times each; and 10 subjects selected at random from the 20, ranked them 10 times each. The rankings by each subject were grouped according to the method employed above and the resulting composite rankings or judgments were correlated with the true order of the weights, and the means of the coefficients obtained were calculated. The mean for those who ranked the weights 2 times was .79, for those who made 5 rankings .92, and for those who made 10 rankings each .97.

The results of these two methods are given in comparison in Table IV. The means at the bottom of the table, which were calculated from the actual coefficients and not from the frequency distributions, give the values of the judgments under the various groupings of rankings by different subjects and of rankings by the same subject. The improvement in the validity of the individual judgment caused by increasing the number of rankings upon which it is based practically parallels that caused by grouping an equal number of rankings by a corresponding number of subjects.

The results obtained in this part of the study give further evidence that the increase in validity secured by grouping into increasingly larger groups is not due to the fact that the judgments were made by different persons. Therefore, we conclude that Gordon's results have been misinterpreted.

TABLE IV

A Comparison of the Effects of Grouping Rankings Made by Different Subjects with the Effects of Grouping Corresponding Numbers of Rankings Made by the Same Subject

Frequencies of Coefficients Showing the Correlation between Rankings by Groups of Different Sizes and the True Order of the Weights. (Gordon's Data)					Frequencies of Coefficients Showing the Correlation between Rankings by the Same Subject Based upon Different Numbers of Single Rankings and the True Order of the Weights. (Our Data)			
Size of Groups				Coefficients	Number of Rankings			
10	5	2	1		1	2	5	10
8	13	13	13	91 to 100	13	17	12	10
2	5	23	37	81 to 90	37	18	7	
	2	6	21	71 to 80	21	5	1	
		4	12	61 to 70	12	2		
		4	7	51 to 60	7	3		
			1	41 to 50	1	1		
			2	31 to 40	2	2		
			2	21 to 30	2	0		
			1	11 to 20	1	1		
			2	1 to 10	2	0		
			0	− 9 to 0	0	0		
			2	−19 to −10	2	1		
.96	.92	.83	.72	Means	.72	.79	.92	.97

4. *The Effects of Grouping 'Chance Judgments'*

A chance order was determined by the order of exposure of well shuffled cards which were numbered from 1 to 10. Two hundred chance order arrangements thus secured were each correlated with the true order. The resulting coefficients ranged from .82 to − .80 with a mean of − .034 (practically 0, as would be expected) and a S.D. of .37. These chance order rankings were grouped and correlated with the true order according to the method employed by Gordon. The means of the coefficients, which indicate the value of the chance judgments are − .070 for the group of 5, − .144 for the group of 10, − .231 for the group of 20, and − .235 for the group of 50. The data are shown in Table V in frequency distributions. The means are calculated from the actual coefficients.

It is clear from the table that there is a slight negative influence over chance, the mean of the correlations of the

TABLE V

FREQUENCIES OF COEFFICIENTS SHOWING THE CORRELATION BETWEEN 'CHANCE JUDGMENTS' IN GROUPS OF DIFFERENT SIZES AND THE TRUE ORDER OF THE WEIGHTS

Coefficients	Size of Groups				
	1	5	10	20	50
80 to 89	2				
70 to 79	3	1			
60 to 69	3	1			
50 to 59	8	0			
40 to 49	6	1	1		
30 to 39	20	3	1		
20 to 29	15	2	2	2	
10 to 19	18	5	2	1	
0 to 9	21	4	3	0	
− 10 to − 1	18	2	1	2	1
− 20 to −11	14	2	2	0	1
− 30 to −21	16	8	0	1	0
− 40 to −31	22	5	4	1	1
− 50 to −41	23	4	3	1	1
− 60 to −51	3	0	1	1	
− 70 to −61	5	1	0	0	
− 80 to −71	3	1	1	0	
− 90 to −81				0	
−100 to −91				1	
Means	−.034	−.070	−.144	−.231	−.235

original rankings with the true order being − .034. We are not particularly concerned with why this mean is negative but with the effect of grouping such data by Gordon's method. The effect of grouping, as indicated by the means at the bottom of the table, is an increase in the numerical value of the coefficients with no change in sign. This is what one would expect. If we mathematically combine measures having any peculiar characteristic that is generally common to the group the peculiar trait will be exaggerated. This is exactly what happens when data are grouped after Gordon's method. If the general trend as indicated by the mean of the original coefficients is negative, grouping gives a larger negative coefficient; if it is positive a larger positive coefficient. The increase is a sort of summation of the cause whatever its nature.

5. *The Effects of Grouping Compared with Predicted Values*

Kelley (1925) applied the Spearman-Brown formula to Gordon's results, after transmuting Rho values to r values. He found the predicted increase in validity very near the actual increase obtained by grouping. Table VI presents the Rho values from Parts I and III of the present study with their equivalent r values, and predicted values based upon the initial r value. The Rho values were transmuted by means of the table given in the appendix of Educational Statistics by Odell.

TABLE VI

COMPARISON OF ACTUAL AND PREDICTED VALUES

	Number of Judgments Grouped	Actual Validities		Predicted Validities	
		Rho Values	r Values	r Values	P.E.
Data from Part I.....	1	.38	.395	.395	
	5	.68	.693	.702	±.042
	10	.85	.861	.865	±.038
	20	.92	.925	.917	±.016
	50	.96	.965	.970	±.007
Data from Part III.....	1	.72	.736	.736	
	2	.83	.842	.848	±.020
	5	.92	.925	.933	±.010
	10	.96	.965	.965	±.005

It is seen from the table that in no case is there a difference between the actual value and the predicted value greater than the probable error of the predicted value which was calculated from the formula given by Shen (1924). When this is considered together with Kelley's findings mentioned above, it seems evident that the Spearman-Brown prediction formula describes well the increase in r which is to be expected when such data are grouped after this method.

In fact, this method proposed by Gordon to demonstrate that group judgments are more reliable than individual judgments very nearly meets the conditions and assumptions basic to the Spearman-Brown formula. Referring to Brown's derivation of the formula (1910), Ruch, Ackerson, and Jackson (1926) say: "In its simple derivation by Brown the only assumptions required are that the standard deviations

of the unitary tests must be equal and that the intercorrelations of all pairs of the unitary tests must be equal." Let us examine Gordon's method for conditions that correspond to these assumptions. Since the true order of the weights contains the numbers regularly from 1 to 10 and each ranking had to make use of the same numbers (no two weights could be given equal rank) the standard deviations will always be equal. This meets the first of the above assumptions. The correlations of the several single rankings (unitary tests) with the true order (the other member of each pair) are not equal. However, the initial r value upon which the prediction is based is the mean of the correlations of the several single rankings with the true order of the weights. And furthermore, the 200 single rankings are grouped according to chance, and the correlations of the group judgments with the true order are averaged. By this system of averaging, the second assumption is approximated. In Lanier's demonstration of the formula (1927) which is similar to that found in Kelley's Statistical Method two assumptions are employed; that the successive measures must have the same means and the same standard deviations. It has been shown already that the standard deviations will be equal, which meets the first assumption used by Lanier. For the same reason the means also will be equal. The numbers indicating positions of weights in the true order and in rankings in every case must be identical, 1 to 10. This complies with the other assumption. Therefore, the assumptions required in Lanier's demonstration are empirically supplied in Gordon's method. Furthermore, the source of Lanier's and Kelley's demonstrations of the formula is Spearman's formula for the correlation of sums (1913). Gordon's method is surely a correlation of sums.

With the facts thus before us, it seems only logical to conclude that Gordon's method does not offer proof of the greater reliability of group judgments but does come nearer than any other experimental 'set up' of demonstrating empirically the validity of the Spearman-Brown prediction formula when the assumptions are realized.

Conclusions

1. If a large number of judgments or rankings of a series of weights (by the same subject or by different subjects) are grouped, correlated with the true order, and averaged after Gordon's plan, the result will be an increase in validity.

2. The increase in validity obtained by grouping data according to Gordon's plan is not due to the fact that the judgments were made by different persons but to the mathematical principle involved in this method of combining data.

3. The actual increase in validity secured by this method of grouping, when ten weights are used, follows closely that predicted by the Spearman-Brown formula.

It seems reasonable that this general conclusion might be deduced in the form of a practical suggestion. Extreme caution should always be exercised in interpreting data which have been treated statistically that the outcomes of statistical manipulation are not mistaken for experimental results.

References

1. Brown, Wm., The correlation of mental abilities, *Brit. J. Psychol.*, 1910, 3, 296–322.
2. Gordon, K., Judgments in esthetics, J. Exper. Psychol., 1923, 6, 36–43.
3. Gordon, K., Group judgments in the field of lifted weights, *ibid.*, 1924, 7, 398–400.
4. Kelley, T. L., Statistical Method, Macmillan, New York, 1923, 196–197.
5. Kelley, T. L., The applicability of the Spearman-Brown formula for the measurement of reliability, *J. Educ. Psychol.*, 1925, 16, 300–303.
6. Ruch, G. M., Ackerman, L., & Jackson, J. D., An empirical study of the Spearman-Brown formula as applied to educational material, *ibid.*, 1926, 17, 309–313.
7. Shen, E., The standard of certain estimated coefficients of correlation, *ibid.*, 1924, 15, 462–465.
8. Spearman, C., The correlation of sums and differences, *Brit. J. Psychol.*, 1913, 5, 417–426.
9. Watson, G. B., Do groups think more efficiently than individuals? *J. Abn. & Soc. Psychol.*, 1928, 23, 328–336.
10. Wheeler, R. H., Readings in Psychology, 1930.

(Manuscript received June 15, 1931)

AN EXPERIMENTAL STUDY OF VIVIDNESS IN LEARNING AND RETENTION

BY W. L. VAN BUSKIRK

University of Pennsylvania

Introduction

The opinion that vividness is a condition favorable to learning enjoys extensive acceptance among psychologists who are concerned with the so-called secondary laws of association. In a critical review of the phenomena of learning, Lashley [1] intimates that intensity and vividness 'apply more widely' than any of the other secondary laws, but he hastens to suggest that this seemingly wide applicability may be due to the fact that "we know so little about them."

In so far as vividness is concerned his point is well taken. The experimental evidence upon which the vividness generalization is based is distinctive in its paucity as well as for its inadequate methodology. All of the few studies are primarily concerned with the relative significance of frequency, recency, primacy, and vividness, and there is little accord as to the importance of the last three. One emerges from a perusal of these works with the conclusion that vividness is a condition favorable to learning of less importance than frequency. A more specific conclusion is impossible.

Moreover, it seems to the writer that the methods used in these investigations were such as to obscure the value of the 'vivid.' In the first place, the material to be learned was exposed only once to a subject. The subject then reproduced by direct recall, prompting, or the treffer method. No one would question the statement that partial learning of this sort does not give a true picture of what occurs in more complete learning. Secondly, methodology is at fault in that there has been no adequate standard for the comparison of the weighted

[1] Lashley, K. S., Nervous mechanisms in learning. *Foundations of experimental psychology*, Clark University Press, Worchester, Mass.; 524–563.

and non-weighted units. The usual procedure is to state the number of times the weighted unit occurred, or to express this in terms of its possible occurrence. Such statements give no reliable clue as to just how much vividness may have affected the occurrence of that unit.

Calkins [2] realized the need for such a standard and constructed one based upon half of her 1200 series. However, she makes little use of it in the interpretation of her results. If one computes the effect of vividness using her standard, it will be found that its average effectiveness of 48 per cent as reported by her dwindles to 21.7 per cent for series of ten and twelve units. Welch and Burnett [3] report the findings of Lamb and Waterman who investigated the relative importance of the four secondary laws. Lamb's standard of comparison consisted of the non-weighted units in the same series. Waterman used the same standard. Recently, Jersild [4] used a standard of the same sort.

The writer believes that such standards are not sufficiently refined to show just what part vividness may have played in the occurrence of the vivid. Comparing the vivid with the average occurrence of all of the unweighted units in the series, fails to account for the fact that the position of a unit in a series influences its occurrence in immediate reproduction.[5] Nor will variation of the position of the vivid in the series overcome the difficulty since it seems likely that the vivid itself may facilitate or inhibit the reproduction of the preceding or following unit or both.

Another consideration which has been neglected in the study of vividness is that of its retentive value. All of the investigations mentioned above are concerned only with immediate reproduction.

The present experiment was devised as an attempt to ob-

[2] Calkins, M. W., Association, *Psychol. Rev.*, 1894, I, 476–483.

[3] Welch, G. B. and Burnett, C. T., Is primacy a factor in association formation, *Amer. Jour. Psychol.*, 1924, **35**, 396–401.

[4] Jersild, Arthur, Primacy, recency, frequency, and vividness, JOURN. EXPER. PSYCHOL., 1929, **12**, 58–70.

[5] Robinson, E. S. and Brown, M. A., Effect of serial position upon memorization, *Amer. Journ. Psychol.*, 1926, **37**, 538–552. See also, Jenkins, J. G. and Dallenback, K. S., The effect of serial position upon recall, *Amer. Journ. Psychol.*, 1927, **38**, 285–291.

tain an objective evaluation of the effect of vividness in learning, and of the persistence of that effect over different intervals of time. It aims to measure this effect in terms of an adequate standard of non-vividness, rather than in terms of the relative significance of primacy, recency, and frequency as compared with vividness.

Such a standard may be constructed by determining the order in which a given subject acquires the units of a series, and from this result selecting a unit as a standard. In a like series, in a position comparable to the position occupied by the "standard" unit in the non-weighted series, is placed a vivid unit. The frequency with which the vivid unit appears in the total trials necessary to learn the series may then be compared with that of the standard unit in the non-weighted series. Such a situation demands the use of nonsense material and necessitates the assumption that the units be of equal associational difficulty. This assumption is probably justified in treating of massed results.

It was decided that the 'standard' position should be that of the unit which occurred with least frequency for any given individual. The choice may be criticised in that there is greater possibility for improvement here than at any other position in the series. However, the choice seems justified in that improvement at this point must overcome the original positional difficulty.

The method has several advantages over the usual procedure. First, by having the subject actually 'learn' the whole series we are able to observe the effect of the 'vivid' throughout the entire process. Secondly the effect of individual variations in positional difficulty is eliminated, since the effect of vividness is calculated for each individual in terms of what, for him, is the most difficult position. Thirdly, the effect that the 'vivid' may exert upon other units of the series does not affect the standard in any way.

The only problem which remains is to satisfy the vivid condition. In this case, the "vivid" is a large red syllable on a green background, as compared with a smaller black syllable on a white ground. Increase in the size of the letters fulfills

the condition of those who would define vividness as the intensity of stimulation. The combination of color and increased size, appearing in an achromatic series, should satisfy those who hold that the "vivid" is the unusual. Again, its color should meet the approval of those who agree with the thought of Calkins [6] who defined vividness as the 'vague and elusive,' the 'unusual,' or the 'object of instinctive, and therefore of psychologically inexplicable interest.' Finally, it adds something to all of this in that red letters on a green background are far less legible than black letters on a white ground.[7]

Conditions

Subjects.—One hundred and two subjects were used in the experiment. Eighty-five were members of the upper classes in the Wharton School of the University of Pennsylvania; two were students in the School of Education in the same institution; and fifteen were pupils in the high school at Blakely, Pennsylvania.[8] Of the total, seven were female and ninety-five were male. The university subjects were volunteers; the high school pupils were sent to the E by the teachers in charge of the study hall.

None of the S's knew the object of the experiment. Most of them believed E to be working on an advertising problem.

Except for a few cases in relearning the experiment was performed with one S at a time. A few of the S's forgot the exact time of their appointments, in which case E permitted them to relearn with one other S present.

Apparatus.—The apparatus consisted of a metronome and three series of nine nonsense syllables, each syllable on a cardboard background, six by ten inches, as follows:

Series I: pef, yib, vor, zut, ras, teg, nad, sov, fik.
Series II: tol, hir, gah, pem, dak, fon, tuv, nim, ker.
Series III: series two in red and green.
In series I and II the letters were 120 point, black, gummed Gothic 'caps' on a white ground. Series three differed in that the letters were 144 point red Gothic 'caps' on a green background.

Most of the syllables were selected from those used by Lamb and Waterman.[9] In placing them in the series the E arranged them so that the same vowel occurred with the same frequency in each series. This was an attempt to carry the equalization of difficulty a step further than that implied in the nature of the nonsense syllable.

---

[6] Calkins, M. W., Association, *Psychol. Rev.*, 1896, 3, 32–49.

[7] Outdoor Advertising—the Modern Marketing Force, Outdoor Advertising Association of America, 1928, 76–78.

[8] The writer wishes to acknowledge his indebtedness to Mr. H. B. Anthony, Superintendent of Schools, for his coöperation in securing these subjects.

[9] Lamb, H. W., Primacy, recency, frequency, and vividness. Waterman, M. O. Primacy, recency, frequency, and vividness. Reported by Welch, G. B. and Burnett, C. T., Is primacy a factor in association formation, *Amer. Journ. Psychol.*, 1924, 35, 396–401.

Method.—S and E were seated at opposite ends of a five foot table when the work was done at the university. At the high school S and E sat facing each other at class room desks placed approximately five feet apart.

In presenting a series E held the cards perpendicular to the surface of the table with his left hand in such a manner that as he pushed one card forward with his right hand, the card behind it was exposed to S. The first syllable in each series was preceded by a blank white card. At E's left a metronome beat 70 times per minute. On every second beat a card was pushed forward.

Learning I

Instructions: "I am going to show you a series of nonsense words. Nonsense words are three letter words which have no meaning. There are nine words in the series, one on each of these cards. I shall show them to you one at a time. When you have seen all the words, write as many as you can remember on the paper. When you have written all of the words that you can remember, cover what you have written with the cardboard and I'll show you the words again. Again, you write all of the words that you remember regardless of whether you have written them before or not. Every time write every word that you remember. This will continue until you can write all nine words after a single showing. The object is to learn the series. Do you understand?"

If S asked if the 'words' were to be learned in order, he was told to learn them as he pleased.

Upon completion of instructions Series I was shown repeatedly to S, one syllable at a time. After each showing S wrote as many of the words as he could remember. When S reported that he had nine syllables, E examined them and told S how many were correct. If additional presentations were needed the experiment continued until S wrote all nine syllables correctly.

S's paper was then examined to find the syllable which occurred least frequently in the total number of trials. In the event of a tie, the syllable nearest the center of the series was selected, unless the results indicated a tendency of the S to acquire the syllables successively from one end of the series to the other. In this case the syllable nearest to the end was chosen.

Learning II

As soon as this calculation was made, Series II was prepared by placing a vivid syllable from Series III in the same position in Series II that the last learned syllable occupied in Series I.

Instructions: "Here is another series of the same kind. There are nine words in it. Will you learn it as you learned the first?"

When Series II had been learned S was asked to report again at a definite time, either one week or two weeks later. E also requested S not to think of what he had learned.

The subjects were divided into two groups. Group I consisted of 51 S's who reported after an interval of one week. Group II, in which there were 50 S's [10], reported after a two week interval. The selection of these groups was determined solely by the days on which the S's appeared for the original learning. Those who reported on Monday and Tuesday were assigned to the two week group. Those who reported on the remaining days were asked to return one week later.

The same method of relearning was used for both groups. First S reproduced as many syllables as he could by direct recall, after which he relearned each series. In

[10] Fifty-one S's were assigned to this group. One failed to report for relearning.

relearning, a black syllable on a white ground was substituted for the red syllable in Series II.

Instructions: "Take this paper and write as many of the words as you remember regardless of the series in which you learned them. Just write them down as they come to your mind."

When S announced that he couldn't remember any more, Series I was relearned.

Instructions: "Now we'll learn Series I again."

Next, Series II was relearned.

"Now we'll learn Series II again."

Results

The effect of vividness is to be measured in terms of the occurrence of a vivid syllable in one series as compared with the occurrence of a standard syllable in another series. Since, only a few S's learned both series with the same number of presentations for each series, it seemed advisable to present the results as percentages. Accordingly, the percent occurrence for the last learned syllable and for the vivid syllable were computed for each S. These, in turn, were averaged and presented in the tables under the heading 'Average Occurrence.' The percentages represent the number of times the S's reproduced the given syllable in relation to the total possible times they might have reproduced it; *i.e.*, the number of presentations of the series in which the syllable occurred. Thus, for any given S, the occurrence of a syllable was recorded as 20 percent if that syllable occurred twice in learning a series for which he required ten presentations.

The results for learning have been presented in two groups to facilitate the reader in comparing the results of each group in learning with their results in relearning.

TABLE I

Learning

Group	Number of Subjects	Standard Average Occurrence (%)	Vivid Average Occurrence (%)
I	51	25.53 (11.36)	82.04 (19.53)
II	51	27.55 (11.78)	84.70 (20.11)
I and II combined	102	26.54	83.37

The Roman numeral under 'Group' indicates whether the S's in the group reported for relearning one or two weeks later. In the column headed 'Number of Subjects' is indicated the

number of subjects comprising the group, which is the basis upon which the averages are computed. Under the column 'Standard' is given the average occurrence of the standard syllable. Under 'Vivid' is shown the average occurrence of the vivid syllable. The figures in parentheses beside each average are the standard deviations of the averages. The bottom row represents the combined learning results for the two groups.

TABLE II

RELEARNING

Group	Number of Subjects	Standard Average Occurrence (%)	Vivid Average Occurrence (%)
I	51	64.69 (27.88)	84.45 (23.22)
II	50	58.38 (25.69)	75.82 (27.80)

The Roman numeral I indicates that the relearning took place after an interval of one week while II indicates a two week interval. The averages are based upon the number of times the vivid or standard was reproduced in direct recall and relearning in proportion to the number of times it might have occurred. If the syllable occurred in direct recall it was recorded as 100 percent; if not, this was recorded as one presentation in which it might possibly have occurred. The direct recalls are included in the averages.

TABLE III

DIRECT RECALL

Group	Standard		Vivid	
	Number of Subjects	Percent of Group	Number of Subjects	Percent of Group
I	15	29.41	33	64.71
II	9	18.00	25	50.00

Under "Standard" and "Vivid" are shown the number of S's who recalled the standard and vivid syllables after one and two week intervals before the material was presented for relearning.

1. Under the heading Av. Vivid-Av. Standard are indicated the conditions and groups for which this difference was

TABLE IV
DIFFERENCES

	Obtained Difference	Sigma Difference	Reliability $\frac{\text{Obt. Diff.}}{\text{Sigma Diff.}}$
1. Av. Vivid-Av. Standard			
A. Learning			
Group I	56.51	3.18	17.77
Group II	57.15	3.27	17.48
B. Relearning			
Group I	19.76	5.08	3.89
Group II	17.44	5.35	3.26
2. Learn V-Relearn V			
Group II (2 wk.)	8.88	4.85	1.85
3. Relearn St'd-Learn St'd			
Group I	39.16	4.23	9.26
Group II	30.83	3.96	7.78

found. The heading indicates that the average occurrence of the standard syllable has been subtracted from the average occurrence of the vivid syllable. Thus, the difference represents the actual effectiveness which may be attributed to vividness. The significance of the difference is indicated in the last column.

2. Under Learn V-Relearn V is indicated the difference between the occurrence of the vivid syllable in learning and its occurrence in relearning after a two week interval.

3. Under Relearn St'd-Learn St'd are shown the increases in relearning of the occurrence of the standard syllable over its occurrence in learning.

DISCUSSION

The results indicate that vividness, as defined in the present investigation, should occupy a firm position as one of the more important of the conditions favoring learning. In Table IV, the difference between the occurrence of the vivid and the standard syllables in learning is high for each group. (Group I, 56.51 percent; group II, 57.15 percent.) These differences reveal that vividness effects a marked economy in learning, and their exceptional statistical significance adds weight to the contention.

It is interesting to compare these figures with those reported in other papers. In this respect the present experiment shows an advantage for the vivid which is far in excess of that shown in previous reports. After subtracting the 26.3

percent occurrence of her standard from Calkins' [11] 48 percent occurrence of the vivid, the present findings exceed hers by 35.12 percent. In a later paper [12] she reports vividness as 52.2 percent against a normal of 20.8 percent. Lamb [13] reports an advantage for the vivid which is only 7 percent in excess of his 36 percent occurrence of all of the unweighted syllables in the series, while Waterman's [14] report is still lower being only 2 percent higher than the 20 percent occurrence of the unweighted units.

The same disparity persists when the average occurrence of the vivid unit alone is considered. Again one finds that the 83.37 percent (Table I) average occurrence of the vivid syllable in this experiment is larger than that reported in previous papers. Calkins' reports of 48 percent and 52.2 percent are 35.37 percent and 31.17 percent lower than the present results. Lamb's result is 40.37 percent lower and Waterman's, 61.37 percent lower. Jersild's [15] report is more difficult to compare because his results were treated in a different manner. He reports that the vivid occurred 176 times in a possible 253, or approximately 70 percent of the time according to the manner of treatment in the present paper. This finding is more in accord with that of the present investigation, but was obtained only with what Jersild terms his 'most effective vivid condition'; *i.e.*, 'Now get this.'

When the positional difficulty which vividness had to overcome in the present experiment is considered, the marked economy effected by the vivid condition becomes increasingly significant. It reveals that vividness is of far more importance than has hitherto been shown. The writer believes that the difference between the present results and those reported in previous investigations is due largely to a difference in

[11] Calkins, M. W., Association, *Psychol. Rev.*, 1894, 1, 476–483.

[12] Calkins, M. W., Association, *Psychol. Rev.*, 1896, 3, 32–49.

[13] Lamb, H. W., Primacy, recency, frequency, and vividness. Reported by Welch, G. B. and Burnett, C. T., Is primacy a factor in association formation, *Amer. Journ. Psychol.*, 1924, 35, 396–401.

[14] Waterman, M. O., Primacy, recency, frequency, and vividness. Reported by Welch, G. B. and Burnett, C. T., Is primacy a factor in association formation, *Amer. Journ. Psychol.*, 1924, 35, 396–401.

[15] Jersild, Arthur, Primacy, recency, frequency, and vividness, JOURN. EXPER. PSYCHOL., 1929, 12, 58–70.

method. In former papers the vivid syllable was recorded as effective only if it occurred after the first and only presentation. Consequently, there was less chance for it to be effective than in the present experiment where the material was presented as many times as were necessary for the subjects to attain a given standard of perfection. That the results obtained under the latter conditions are so much higher than those obtained with the former procedure serves to emphasize the point that the present experimental conditions afford a better opportunity for the investigation of the effect of vividness in learning.

As for its retentive value, further inspection of Table IV (Difference) well suffice to show that here, as well as in learning, vividness is a favorable condition. The differences between the occurrence of the standard and vivid syllables both for the one week and for the two week intervals, 19.76 percent and 17.44 percent respectively, lie within the scope of perfect reliability. Table III shows that these differences are considerably larger if limited to direct recall. After an interval of one week 64.71 percent of the subjects recalled the vivid syllable before the series was presented for relearning while only 29.41 percent of them recalled the standard, a difference of 35.3 percent. Likewise, after two weeks 50 percent recalled the vivid as against 18 percent for the standard, a difference of 32 percent.

Comparison of these figures with the differences shown in learning (56.51 percent) seems to indicate that vividness, while clearly effective in retention, is less effective here than in learning. Actually this is not the case. What happens is that the standard shows relative gain. In Table IV, under the heading 'Relearning St'd—Learning St'd,' is found a 39.16 percent gain in the occurrence of the standard after one week and a 30.83 percent gain after a two week interval. The corresponding figures for the vivid syllable may be computed from Tables I and II. We find that after a one week interval the occurrence of the vivid syllable increases by 2.41 percent over its occurrence in learning and that after a two week interval it decreases by 8.88 percent. The differences between the occurrence of the standard in learning and re-

learning for both intervals are much greater than the corresponding differences in the occurrence of the vivid syllable. Thus it appears that that which is difficult to learn (the standard is the last learned syllable in the first series) shows relative gain in retentive value over that which is easily learned (the vivid). This statement receives additional substantiation from the remarks of many subjects upon recalling or seeing the standard syllable in relearning. Often a subject remarked that this was the word which gave him so much trouble in learning.

That vividness is actually as significant in retention as in learning seems to be evident in view of the fact that in relearning after one week it occurred 84.45 percent of the time that it might have occurred. This figure is slightly higher than that for its occurrence in learning. After a two week interval, one finds that its effect decreases by 8.88 percent. The index of reliability (1.85), however, indicates that the decrease is not significant.

Furthermore, it seems unlikely that variations in time alter appreciably the effect of the vivid in relation to the standard since the difference between the occurrence of the vivid and the standard syllables after one week (19.76 percent) is only slightly more than after two weeks (17.44 percent). Inspection of the sigmas of these differences indicates clearly that the difference between them is negligible.

Conclusions

The results of the present study seem to warrant the following statements:

1. The procedure followed in this experiment is more conducive to the investigation of the effect of vividness in learning than the method used in previous studies.
2. The economy effected by vividness in learning is greater than has heretofore been shown.
3. The effect of vividness in learning tends to persist with the passage of time.
4. That which is difficult to learn shows relative gain over that which is easily learned.

(Manuscript received March 23, 1932)

CHROMATIC SENSATIONS FROM FIELD-SIZE MEASUREMENTS AT VARYING WAVELENGTH

BY CHARLES A. RINDE

College of the Pacific

Although investigations of the retinal color-sense mechanism were sufficiently extensive to become the bases of a theory by Thomas Young as early as 1801, even before his formulation of the wave theory of light, the present knowledge of color vision is less satisfactory than that in other branches of optics. Numerous theories, for the most part irreconcilable with one another, exist; and three of these are widely taught.

The oldest, the Young-Helmholtz theory, requires three primary chromatic sensations, each of which is assumed to be produced upon the stimulation of the corresponding receptor mechanism. Working in Helmholtz's laboratory, König showed that all the color sensations can be reproduced by suitable mixtures of specific values of red, green, and violet-blue.

But psychologists have been unwilling to accept the implication that yellow is not visually primary. The majority, therefore, favor a four-color theory, such as that of Hering or Ladd-Franklin. These theories require two pairs of primary chromatic sensations which are values of red-and-green and yellow-and-blue. Hering assumed that the active substance for each pair is continually undergoing assimulative and dissimulative physiological changes in dynamic equilibrium. The effect of light in shifting the equilibrium is supposed to be responsible for the sensation. The shorter wavelength member of each pair is supposed to increase the assimulation in its receptor, and the longer wavelength member, the dissimulation. Dr. Ladd-Franklin supposed that the receptor for each pair contains a single complex molecule capable of breaking down in two different ways.

For example, the red-green molecule would break down to produce either red or green, depending upon the exitation.

It occurred to the writer in 1927 that useful information might be gained by measuring the retinal areas sensitive to various chromatic stimuli. Preliminary to undertaking measurements with spectral stimuli, it was decided to investigate the effects of pigment colors. The following year, therefore, a series of perimetric measurements, was begun under the author's supervision by Mr. C. J. Looser, an optometrist, who used both patients and volunteers as subjects. In the course of two years, 773 eyes were measured with red, green, and blue pigmented targets. Since each field was measured along eight meridians, this gave 6, 184 points for each color. The average fields for each eye were plotted. The result for the right eyes is shown in Fig. 1. Apparently, the areas for the red and green sensations are no more nearly congruent

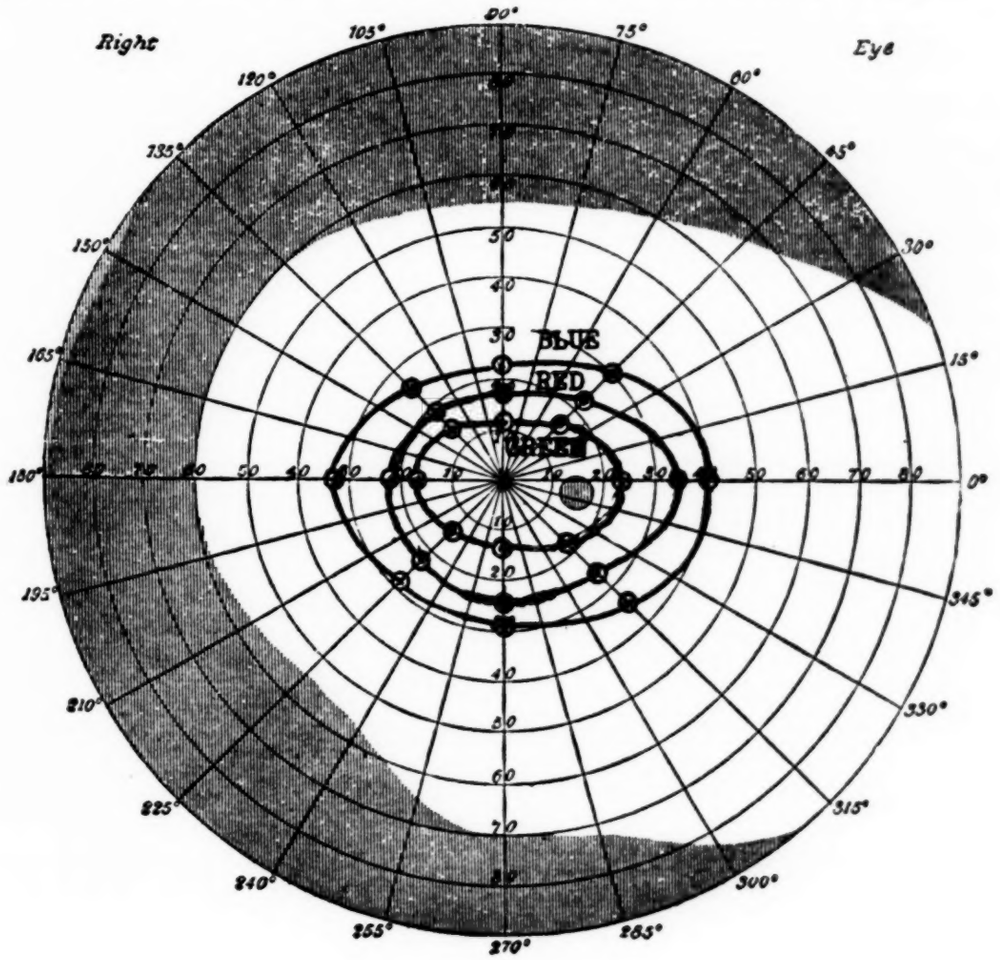


FIG. 1.

than are the areas for the red and blue. These results are in good agreement with those of Abney.[1] But since the stimuli used did not produce simple sensations, no conclusion could be drawn from the non-coincidence of the fields. For example, the green target was frequently reported yellow while outside the green sensation area, and experiments were done with several pigments in an unsuccessful attempt to eliminate this effect, the significance of which was not understood at the time.

To get away from the use of pigments and to facilitate variation of the wavelength of the stimulus, the present apparatus was constructed. Illumination is obtained from a slit, collimator, and grating system, with a carbon arc source. From the resulting spectrum, about 31 cms in length, a band 2 mm wide was selected and allowed to fall upon a translucent glass disc 6 mm in diameter. The wavelength was varied throughout the visual range according to an arbitrary scale, which was eventually calibrated in wavelength by the spectrometer method. This whole apparatus was rotatable aboul a vertical axis which was marked by a centering plate designed to give the observer's eye the correct position. The illuminated target thus would describe an arc of 42 cm radius, with the subject's eye as its center. At the mid-point of the arc was the stationary fixation-point, which consisted of a faintly-illuminated gray ground-glass disc of 3 mm diameter.

In making the measurements the subject was seated in a darkened cylindrical enclosure, the interior being painted a neutral gray, with only the fixation-point visible. While the eye remained fixated, the illuminated target was moved steadily into the field of view. At the first appearance of each color sensation, the angle to the left or right of the fixation-point was read to the nearest tenth of a degree. This angle is called 'field-size,' although it is merely the horizontal extent. The values, except in easily-recognized instances of faulty fixation, were usually consistent within a range of approximately one degree. Each point was determined by the average of three or more readings.

[1] Abney, Sir Wm. de W., Researches in Colour Vision, London, 1913, pp. 195–198.

In order to reduce the probability of suggestion and distraction, the enclosure was designed to keep the subject from seeing either the operator or the apparatus, and there were no marks of any sort by which the subject could duplicate positions of the target. A ventilating fan which ran continuously greatly restricted hearing. The subjects were not instructed as to what color they were to see, although they were sometimes permitted to look at the target directly with both eyes before the observations began. When this was done, they were warned that the color they would first observe might or might not be the same as they had just seen. The prejudicial effect of this preview of the color was found to be negligible, except in a very few cases. None of the subjects were acquainted with visual theory or the possible significance of the measurements.

Preliminary measurements were made with an arbitrary wavelength for each color, chosen to give what most observers agreed upon as representative red, green, yellow, and blue. The fields for each eye were paired in six combinations and the size differences obtained. The average differences, taken regardless of sign, are shown in Table I. It is to be

TABLE I

Color Pair	Average Difference in Field Sizes (Degrees)	No. of Pairs Measured
*Red-Green	3.2	70
Red-Blue	4.2	67
Red-Yellow	4.2	49
Green-Yellow	5.7	51
Green-Blue	5.3	68
*Yellow-Blue	4.3	41

noted that the starred pairs do not show an appreciably closer size relationship than do the others.

These data fail to suggest any paired relationships for the chromatic sensation; but there are two major reasons for not concluding that the sensations are unpaired. First, it is known that the field size is a function of both intensity [2] and

[2] Aubert, Physiologie der Netzhaut, 1865, p. 116.

Landolt and Snellen, Handbuch der ges. Augenheilkunde von Graefe und Saemische, 1874, 3, p. 70.

Charpentier, Augustin, La Lumière et les Couleurs, Paris, 1888.

Ferree, C. E. and Rand, G., Chromatic Thresholds of Sensation from the Center to the Periphery of the Retina, Part I, *Psychol. Rev.*, 26, 1919, pp. 16–41.

saturation.[3] Secondly, the wavelengths for the four stimuli were chosen arbitrarily without any reason for supposing their effects to be psychologically or physiologically simple.

Experimental evidence of complexity appeared in the tendency of both the red and the green to produce the sensation yellow in the outer region of the retina. In the green, shortening the wavelength failed to eliminate the yellow response until blue appeared. In the range 480–490mμ it was found that yellow was first produced, then a nondescript bluish gray, and finally, very near the center of the retina, a bluish green. This effect was explained by assuming that the light first stimulated the receptor for yellow; then in a smaller area, that for blue also, the nondescript bluish gray resulting; and in the central region it excited the green in addition to the other two. These effects were all discovered by the subjects entirely without previous suggestion.

The next step was to measure the variation of field size with wavelength. This was done for wavelengths above 445 mμ, giving curves resembling I, II, III, and part of IV in Fig. 2. For each of the four sensations there appears a

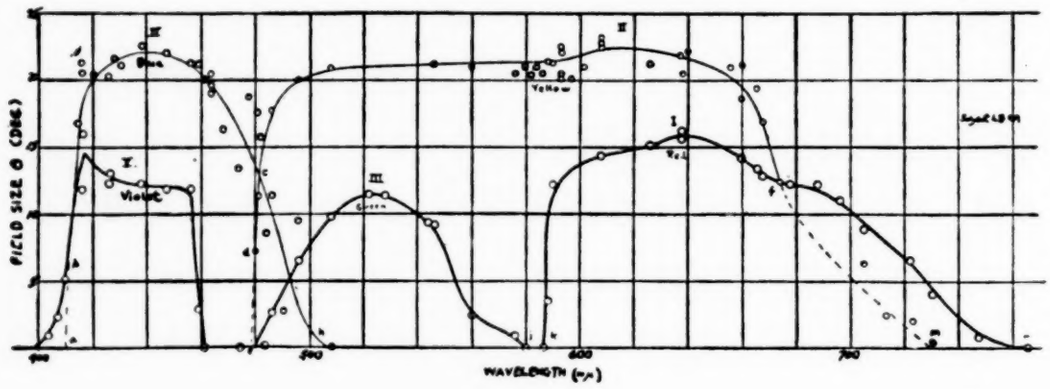


FIG. 2.

fairly definite wavelength corresponding to the maximum area. Varying the wavelength away from this value in either direction decreases the area.

This effect may be ascribed either to a progressively

Ferree and Rand, The Limits of Color Sensitivity: Effects of Brightness of Preexposure and Surrounding Field, *Psychol. Rev.*, **27**, 1920, pp. 377–398.

[3] Bull, Ole, Studien über Lichtsin und Farbensinn, Albrecht von Graefe's *Archiv. f. Ophth.*, **27**, 1881, pp. 54–154.

greater median sensitivity [4] of the receptor field units toward the center of each field, or to a more compact arrangement of the individual receptors near the center. In either case, the greater the effective stimulation, the larger would be the area giving the sensation. The apparent fact that a receptor mechanism absorbs a maximum of energy at some specific wavelength, and absorbs less above and below that value can be explained by an electrical analogy. If an electric circuit has a certain value of inductance and capacity, it will absorb energy in the form of electrical oscillations of a definite frequency provided such energy is available. Suppose an oscillator to be producing electromagnetic waves of the given frequency, the resonant circuit will absorb a maximum of energy. But if the frequency of the oscillator is increased or decreased, the energy in the resonant circuit will fall off rapidly, the curve, except for harmonic effects, having the shape of a rather compressed probability curve.

It must be emphasized that the measurement was not of *intensity*, but of the field-size within which homogeneous light produced chromatic sensations. Further, it should be mentioned that the values for the curves were measured horizontally, on the temporal side of the retina.

It seems significant that in all areas where curves did not overlap, a sensibly constant hue was reported throughout the range of a single curve; while in regions under two or more curves, the resulting sensation was a function of both λ and θ. This suggests the possibility that each curve represents approximately the wavelength range stimulating a single mechanism giving a specific sensation. The hues observed in areas under more than one curve would then be blends depending for their values upon the relative intensities of stimulation of the respective mechanisms.

The foregoing, however, fails to explain the bluish violet hues. For only the one mechanism, that for blue appears to function at wavelengths below about 460 mμ. This would lead to the expectation that only the sensation blue would be

[4] *Cf.* Selig Hecht, The Relation between Visual Acuity and Illumination, *J. Gen. Physiol.*, 11, 1928, pp. 255–281.

observed in the lower end of the spectrum. It is also to be seen that there is a decided asymmetry in the manner in which the curves cover the visual range. These considerations, together with the fact that several observers consistently reported a well-defined inner field, led to further investigation in the shorter wavelengths. The result was a fifth curve for a sensation which I shall tentatively call 'violet.'

Curve V lends itself well to two interpretations. It may be considered as giving the limits of a uniquely short-wavelength sensation, which, when combined with blue, produces the effects observed in the lower part of the spectrum. On the other hand, the curve may indicate the presence of a second range of sensitivity for the red, so that the effect in the short wavelengths is nothing more than reddish-blue.

The latter view is most consistent with the literature; [5] although the first interpretation would not prevent the production of the reddish-blue sensation in any case where the stimulus has wavelength components lying within both ranges I and IV: The possibility of a violet distinct from purple was expressly denied by Dr. Ladd-Franklin.[6]

These curves can be reconciled with the notion of four stimuli producing invariable hues from the center to the periphery of the retina.[7] For the range 460–480 gives 'in-

[5] H. v. Helmholtz, Physiologischen Optik, 2d Ed., Leipzig, Voss, 1896, p. 471.

G. Révész, Über die vom Weiss ausgehende Schwächung der Wirksamkeit farbiger Lichtreize, *Ztsch. f. Psychol. u. Physiol. der Sinnesorgane*, **41**, 2, 1907, pp. 102–118.

Abney, *op. cit.*, 1913, pp. 255–257.

E. Schrödinger, Grundlinien einer Theorie der Farbenmetrik im Tagessehen, Section 9, Die Farbtonanderung beim Zumischen von Weiss, *Ann. d. Physik* (4), **63**, 1920, pp. 512–517.

F. W. Edridge-Green, The Change of Hue Caused by the Addition of White Light to Spectral Colors, Report Brit. Assoc. for Advancement of Science, 89th Meeting, Edinburgh, 1921, p. 447.

D. B. Judd, The Empiric Relation between Dominant Wavelength and Purity, *Jour. Opt. Soc. Am.*, **14**, 1927, p. 475.

[6] The Physicist and the Facts of Color, *Science*, **66**, 1927, p. 591.

[7] Bull, *loc. cit.*

Hess, Carl, Ueber den Farbensinn bei indirectem Sehen, Albrecht v. Graefe's *Arch. f. Ophth.*, **35**, 4, 1889, pp. 17–25.

Hegg, Emil, Zur Farbenpenmetrie, Albrecht v. Graefe's *Arch. f. Ophth.*, 1892, **38**, 3, pp. 145–168.

Baird, J. W., The Color Sensitivity of the Peripheral Retina, Carnegie Institution, Washington, 1905, pp. 59–73.

Dreher, E., Ztch. f. Psychologie u. Physiologie der Sinnesorgane, **46**, 2, 1911, pp. 2–82.

variable blue,' and the range *jk* gives 'invariable yellow,' while any red with sufficient blue added to cancel the effect of peripheral yellow gives 'invariable red,' and similarly a stimulus producing green and sufficient blue to cancel the yellow is 'invariable green.'

The dotted portions are those which have not been satisfactorily determined. The segment *ab* would have to be found by illuminating the target at wavelengths in the vicinity of 410 mμ, and noting, first the violet, and then a slight shift of the sensation toward the blue. This was apparently done in two cases, but not with sufficient consistency to justify a final curve. Segment *de* should indicate the change from a blue to a slightly bluish gray, just as does segment *dc*, but the effect was too weak to be observed with certainty. Segment *ci* was determined by a shift from yellow to a bluish gray sensation, and *ih* by a change from green to bluish green. The portion of the yellow curve, *fm*, was not ordinarily determinable, apparently because the intensity of the red sensation was so great as to mask any shift toward the yellow (*i.e.* orange). The points shown were found in one instance only, when a subject having yellow and red curves very similar to those of the figure was remeasured after an attack of influenza. The red was greatly reduced in extent, but the yellow was practically unaltered throughout its whole range. The result was that the segments *fm* and *fn* coincided temporarily, so that *fm* represented the beginning of the combined sensation of orange. Unfortunately, such a circumstance has arisen but once, so the result could not be verified. The range *jk* was checked repeatedly at very close values of λ, and so far, curves I and III have never touched or overlapped, but there is no reason for supposing that they would not do so with brighter stimuli. It has not been proved that only five curves are possible. But, so far, attempts to break up those shown have not succeeded. For instance, if one tries to plot a separate orange, it will be continuous with the red at its upper end. This fact obviates any subjective difficulties with color names.

In an effort to verify the relation of field-size to the

intensity of the stimulus, a set of curves was made, at the suggestion of Dr. M. Luckiesh,[8] using a filter of didymium glass between the grating and the target. The result is shown in Fig. 3. The field-size curves *R*, *Y*, and *G* are for the

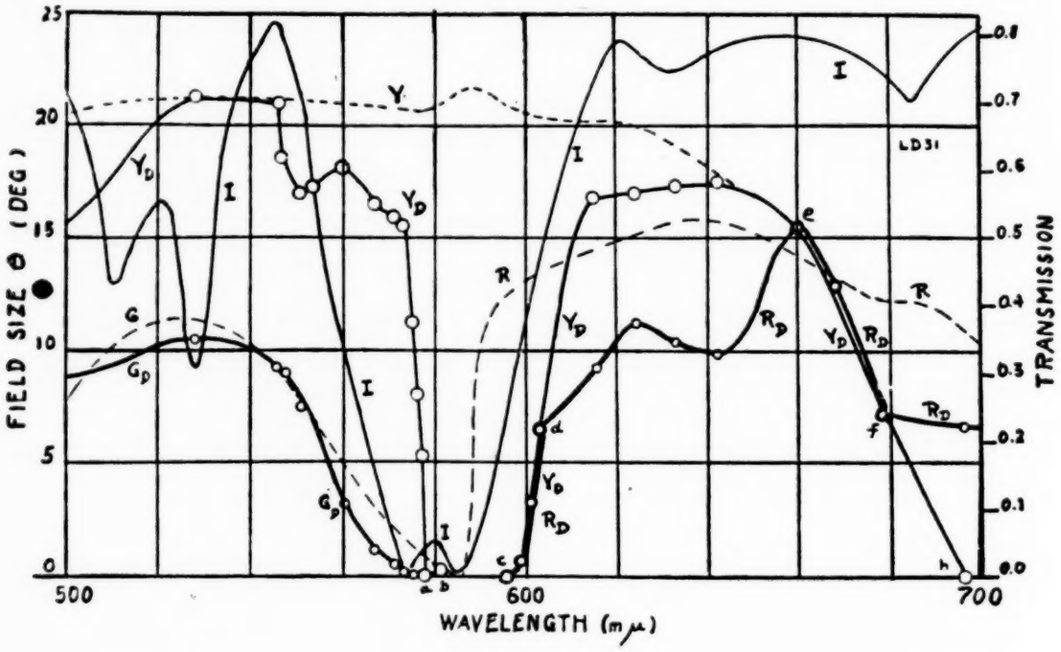


FIG. 3.

sensations red, yellow, and green, as measured without a filter. The curves labelled with the subscript *D* were taken with didymium glass filter in place. All other conditions were as nearly the same as possible for the two sets. No chromatic sensation whatever appeared between *a* and *c* except in the vicinity of *b*, where there was no more than a trace. Along *cd* and *ef* the yellow and red curves coincide, the sensation being orange from the start. The segment *fh* is doubtful, having been determined by the point *h*, the first at which it was found that no traces of yellow could be detected.

The glass used was made by the Corning Glass Works; and in the absence of equipment suitable for measuring its transmission, it was assumed to have substantially the characteristics indicated by the curve I.[9] But it must be re-

[8] Private conversation at Nela Park.

[9] The Ultra-Violet and Visible Transmission of Various Colored Glasses, Gibson, Tyndall, and McNicholas, Bureau of Standards Technologic Papers, 1920, 148, p. 9.

membered that I does not represent the actual intensity curve for the illumination, for no account has been taken of the distribution of energy of the arc. It also appears that all the field-size curves taken with the filter should be shifted about 5 mμ toward the shorter wavelengths in order to fit satisfactorily the yellow absorption band in curve I. The cause of this has not been determined, but is probably due to a displacement of the apparatus during the insertion of the glass plate. However, there is sufficient correspondence between I and the field-size curves to show a relationship between field-size and intensity.

Finally, here is a clear verification of the production of the yellow sensation by wavelengths other than those usually thought of as 'yellow.' For not only were these wavelengths eliminated from both the grating spectrum and any effective extraneous light, but it was also shown that in the absorption region no recognizable chromatic sensation was observed. This made it quite impossible that the sensations observed at other wavelengths could have been stimulated by the so-called 'yellow' light.

Summary of Conclusions

1. In general, homogeneous light does not produce simple sensations. The only exceptions appear to be in the ranges λ 460–480 mμ and 580–585 mμ.

2. The chromatic sensation produced by the violet end of the spectrum is a compound of blue, either with red or with an additional short-wavelength sensation.

3. It is believed that the complementary cancelling effect of blue and yellow sensations produced by a homogeneous stimulus has been found in the range 480–495 mμ.

4. Only five curves, greatly overlapping, have been found which do not become continuous.

5. The limits and the sensitivity ranges vary widely with stimulus intensity. The present paper deals with a moderately low intensity.

The number of satisfactory subjects was limited both by the length of time required for the measurement of a full set

STATISTICAL DATA

Type of Measurement	Number Measured		No. Substantially Agreeing with Foregoing Results. (Eyes)
	Eyes	Subjects	
Field Sets for Red, Green, Yellow, and Blue.....	71	36	62
Field Sets with Violet Included.....	18	10	15
Complete Sets to Check the Five Curves.....	10	5	10
Complete Sets with Didymium Glass Filter.....	4	2	4

Number of Individual Field Measurements with Monochromatic Light
Red 776 Yellow 1,265 Green 788 Blue 722 Violet 383
Total giving positive results 3,934

of fields and by the tendency of subjects to become fatigued and unreliable before completion. Since visual traits are subject to individual variation, the numbers do not justify the assumption that the curves of Fig. 2 are generally characteristic in shape and extent, even for the stimulus intensity used. But should the curves be proved to represent the functioning of the chromatic receptor mechanisms, it seems likely that the number of curves may be significant. For numbers as small as five, the fingers and toes, for example, are likely to present relatively few individuals differing from the mode. Two subjects found to be suffering from color-sense defects have not been included in the discussion.

In order that the experiment should be conclusive, the fields should be measured on at least six meridians besides the horizontal, and intensity of illumination should be controlled so that the effects of changes in brightness might be measured.

The writer's genuine gratitude is due Dr. Samuel R. Cook for his thorough coöperation in every phase of this work. Thanks are also due Dr. M. Luckiesh, Prof. T. D. Phillips, and Dr. D. B. Judd for their valuable suggestions and criticisms.

(Manuscript received July 7, 1931)

EFFECT OF BACKGROUND ON VISUAL ACUITY OF CIRCLE GRIDS

BY EDWARD B. GREENE

Psychological Laboratory, University of Michigan

Problem.—In the measurement of visual acuity one of the problems involved is the effect of the immediate background upon resolution when the test object itself is held constant. Since the problem when applied to circle grids does not appear to have been specifically and numerically answered in any literature available to the writer, the following experimental approach was made as part of the development of a non-language visual acuity chart.

Materials.—1. The test objects were circles with grids of black parallel lines printed on dull finish white photographic paper. The black lines had a reflection factor based on calcium light of 13.2 percent and the white of 68.3 percent. The lines and spaces of the grids were both 1.74 mm wide, subtending an angle of approximately one minute at six meters. Two sizes of circles were used: Series 1, 2.78 cm in diameter, and Series 2, .870 cm in diameter. Hence the circles in Series 1 subtended visual angles of 16″, and in Series 2, 5″.

2. The backgrounds for these test objects were four dull finish paper discs, 8 cm in diameter. The reflection factors based on calcium light were as follows: White, 66.2 percent, Light Gray, 37.7 percent, Dark Gray, 19.5 percent, and Black, 12.1 percent.

3. Exposure Apparatus: A metal box was constructed to control wavelength and intensity of illumination. Frosted Mazda bulbs were mounted on either side of the test object, which was held in a vertical plane, so that the light passing through a flashed opal screen fell on the test object and its background at an angle of 45°. This produced 9.6 candles

per square foot brightness on the background in the test box which was held constant throughout the experiment. The room illumination from similar lamps reflected from cream colored side walls was approximately 4.4 foot-candles. There was no glare.

The test object was mounted on an axle which could be rotated by hand to any desired visual axis.

A timing shutter operated by a pendulum exposed test objects for one second. A light gray spot on the shutter just in front of the test object was used for fixation. This spot had a reflection factor of 37.7 percent and was illuminated to nearly the same brightness as test object.

Subjects.—Six college men were selected for nearly normal vision by retinoscopic procedure and astigmatic charts at six meters.

Procedure

1. *Preliminary Experiment.*—Subjects were seated comfortably in a chair at the distance of seven meters from the test object. The left eye was obstructed with a gray soft paper disc placed in a light trial frame. The subject then reported verbally the direction of the grid line on a white background by naming the time on a clock face which corresponded to the direction of the grid line. Twenty-four readings (hereafter referred to as the standard block) were made, using the six axes of a clock face, four times each in random order, with a one second exposure, and approximately a three second interval. The word 'Ready' was given orally approximately one second before the test object appeared. The standard block was limited to twenty-four readings because this seemed to be near the maximum number which could be shown at this rate without producing a little discomfort due to eye strain and fatigue. To avoid memory aids, the random order of axes was changed on each repetition of a block.

The results of Series I showed that three of the subjects resolved all of the positions of the grid, two subjects resolved with an average of 75 percent correct, and one subject with an average of 65 percent correct. In order to equate the subjects

for efficiency of resolution, each was placed at a distance from the test object where he resolved four standard series with an average between 50 percent and 60 percent correct. The extreme variation of distance between any two subjects was 1.2 meters, and the average distance from the test object was 7.3 meters.

The experimenter encouraged persons to sit at ease and make a reasonable effort which would not lead to great fatigue, and this seemed to be accomplished. The subjects gave approximately eighteen hours for this work in one hour periods between 2 and 4 P.M. on Mondays and Fridays. They understood the problem and the procedure and willingly coöperated as part of the work of an elementary psychology course. Although not thoroughly trained observers, they gave stable and intelligent coöperation. In addition to the statistical results, the subjects ranked the backgrounds after the final day of work in order of their apparent aid to resolution.

2. *Main Experiment.*—With the efficiency of resolution held nearly equal for all subjects in the manner just described, the standard block of exposures was followed on each of the backgrounds for each subject on four different days. The order of presenting the various backgrounds was varied so that each fell in a different temporal position each day, and the adjacent positions were always differently occupied, as follows:

1.	W	L	D	B
2.	B	D	L	W
3.	L	B	W	D
4.	D	W	B	L

W—White
L—Light Gray
D—Dark Gray
B—Black

Similar preliminary and main experiments were made with Series II, the smaller test object.

Results.—Records were taken by writing the judgment given by the subject on a record sheet next to the number representing the true axis of the exposed grid circle. From this indications may be had of both the errors for each axis, and the amounts of error in degrees.

The errors for each axis, although not bearing directly on the problem of background, have been studied to find out what tendencies were present. Since the subjects were selected because they had nearly normal vision, they showed only slight variations. All subjects resolved the vertical and horizontal axes about equally well, and better than the oblique axes. Table I shows these results for each axis on

TABLE I

RESOLUTION OF VARIOUS AXES. SERIES I. RIGHT EYE

Clock Hour	Axis Degrees	Judgments			Standard Error of %
		Total	Correct		
			No.	%	
1	120	384	227	59.4	2.54
2	150	384	246	64.2	2.50
3	180	384	257	67.0	2.44
4	30	384	231	60.0	2.52
5	60	384	234	61.0	2.52
6	90	384	253	66.0	2.44

Series I in the main experiment. Using Yule's (2) formula for the standard error of a proportion, $\sqrt{\frac{PQ}{N}}$, it follows that any difference of more than 10.0 between the percents correct is reliable, to the extent of three times the standard error of the difference. Inspection of the table shows that the difference between the horizontal and vertical axes is 1 percent (67 percent − 66 percent) or about one third of its own standard error. It will doubtless be wise to consider this difference of 1 percent of small importance because of the nature of the variables, for three of the subjects actually did slightly better on the vertical than the horizontal axis. However, we are probably safe in concluding that there is a fairly reliable difference between the oblique axes and the horizontal or vertical, for all subjects show this tendency, and it agrees with the very careful work done by Weymouth (1).

The amounts of errors, disregarding the axis, have been

recorded in Table II, under five classes: correct, 30° errors, 60° errors, 90° errors, and no answer. This shows a tendency in both series for maximum resolution on the Light Gray disc, nearly maximum resolution on the Dark Gray, somewhat less on the Black, and minimum resolution on the white background. From Yule's standard error of proportion formula any difference greater than 010.0 is three times its own sigma. Hence the difference between White and Light Gray 11.0

TABLE II

RESOLUTION ON FOUR BACKGROUNDS. AVERAGE OF SIX SUBJECTS. RIGHT EYE

Percent of Total Judgments Falling in Each Class

	Series I. Larger Circles				Series II. Smaller Circles			
	W	LG	DG	B	W	LG	DG	B
Correct	57.3	68.3	66.7	63.2	52.3	59.3	57.9	56.3
30 Degree Error	12.3	12.8	14.3	11.6	13.1	14.5	16.2	15.9
60 Degree Error	8.1	6.2	5.9	5.7	9.7	7.2	9.6	7.1
90 Degree Error	6.3	5.2	2.8	8.6	7.8	6.4	4.2	9.3
No answer	16.0	7.5	10.3	10.9	17.1	12.6	12.1	11.4
Total	100%	100%	100%	100%	100%	100%	100%	100%
Number	576	576	576	576	576	576	576	576

(68.3 − 57.3) is reliable in Series I, but the other differences are all smaller.

The explanation of these results is probably to be found in the contrast phenomena of observation. The background had its minimum contrast effect in the case of the Light Gray disc, and its maximum in the case of the White. The subjective reports of the subjects confirm the results given to the extent that all preferred one of the grays to either white or black. Four subjects felt that the Light Gray was better, and two, the Dark Gray.

Table II also shows that all of the subjects resolved Series I better than Series II and this tendency is reliable according to Yule's formula. The reason for this is probably the difference in energy between the small and large circle. The ratio of the white areas of these two circles is approximately 1 to 9.45, and hence the retinal photolytic effects would be quite different. This effect is somewhat counter-

acted by the fact that the small circles would be resolved through the center of the fovea where acuity is greatest, while the larger circle fell in part on less sensitive areas, although still well within the limits of the macula lutea.

References

1. Weymouth, F. W. et al., Visual acuity within the area centralis and its relation to eye movements and fixation, *Am. J. Ophth.*, 1928, **11**, 947–960.
2. Yule, G. V., Introduction to the Theory of Statistics, 1924. Lippincott, p. 259, formula 2.

(Manuscript received June 29, 1931)

A NEW TYPE OF PSYCHOGALVANOMETER

BY THOS. H. HOWELLS

University of Colorado

There is apparently a considerable need at the present time for a simple, inexpensive psychogalvanometer which is sensitive, stable, easily controlled, and capable of adaptation for various types of experimentation. It is believed that the apparatus which is described below fulfills these demands to a very considerable extent.

The functioning of this instrument involves two distinct operations. First of all a relatively constant current is passed through the tissues of the subject between the points of contact of the instrument, and second, the variations in electrical potential, which are produced at these points of contact as a result of the variations in resistance of the tissues, are measured by a very sensitive voltmeter. The variations in deflection of this voltmeter constitute a measure of the psychogalvanic phenomena.

The maintenance of a relatively constant electrical current through any conductor demands only that the changes in resistance of this conductor be of a relatively small value as compared with the total resistance of the circuit in which it is placed. Under these circumstances a slight increase in the resistance of the conductor, which is in this instance the tissue of the subject, results in a corresponding increase in the EMF at the points of contact, while a decrease in resistance produces a drop in the terminal voltage. Since the psychogalvanic effect is always, in last analysis, the result of variations in resistance of the tissues, it is evident that an accurate measurement of these changes in EMF should constitute a direct measure of the reflex. It is evident that when the psychogalvanic phenomena are registered in terms of changes in current, as in the ordinary type of instrument, these variations are diminished by the resistance external to

the body of the subject. The vacuum-tube voltmeter, which is extremely sensitive and consumes practically no current in operation, seems to be the ideal instrument for direct measurement.

A means of maintaining a direct current of constant magnitude through the tissues of the body has previously been described by the writer. In brief the construction of the psychogalvanometer, which is the subject of this paper, involves merely the connection of a vacuum-tube voltmeter to the electrodes of the stimulus-apparatus which is described in this article.[1] The reading of this article may contribute to a better understanding of the way in which the constant current is produced, measured and maintained.

Figure 1 presents a diagram of the complete apparatus which can be used as (1) an electrical stimulus-apparatus of the type described in the previously mentioned article, as (2) a psychogalvanometer of the direct current type, or (3) a psychogalvanometer of the alternating current type.

The apparatus which is diagrammed within the dotted lines is practically the same as that described in the note referred to, except that a type 227 vacuum tube is substituted and a switch, *S*2, is provided in order to stop the rectifying action of the tube whenever desired. An alternating potential of quite a high value (around 1000 v.) is obtained from the secondary of a transformer, *T*1, which has a secondary of quite high resistance, such as a radio transmitting transformer or a large induction coil with the vibrator shorted. The exact value of this potential is determined by the output voltage of the supply transformer, *T*2, and by the adjustment of a control rheostat, *R*1, in series with the primary of *T*1.[2] This potential is rectified by the vacuum tube, *VT*1, passed through a load resistance (*R*2, about 1 megohm) which serves to increase the relative amount of resistance outside of the tissues. The current is next passed through the tissues of the subject, between the electrodes, *E*, and finally

[1] Howells, T. H., An electrical stimulus-apparatus, *Am. J. of Psychol.*, XLIII, 1931, 122–123.

[2] The output of *T*2 is ordinarily from 1–10 volts and a 0–100 ohm rheostat is preferable.

measured by the meter, $M1$, which is connected in series. This meter is preferably a sensitive galvanometer with a shunt of variable resistance ($R5$, about 200 ohms maximum) which serves to control its sensitivity.

It will be noted that when the switch, $S1$, is closed the

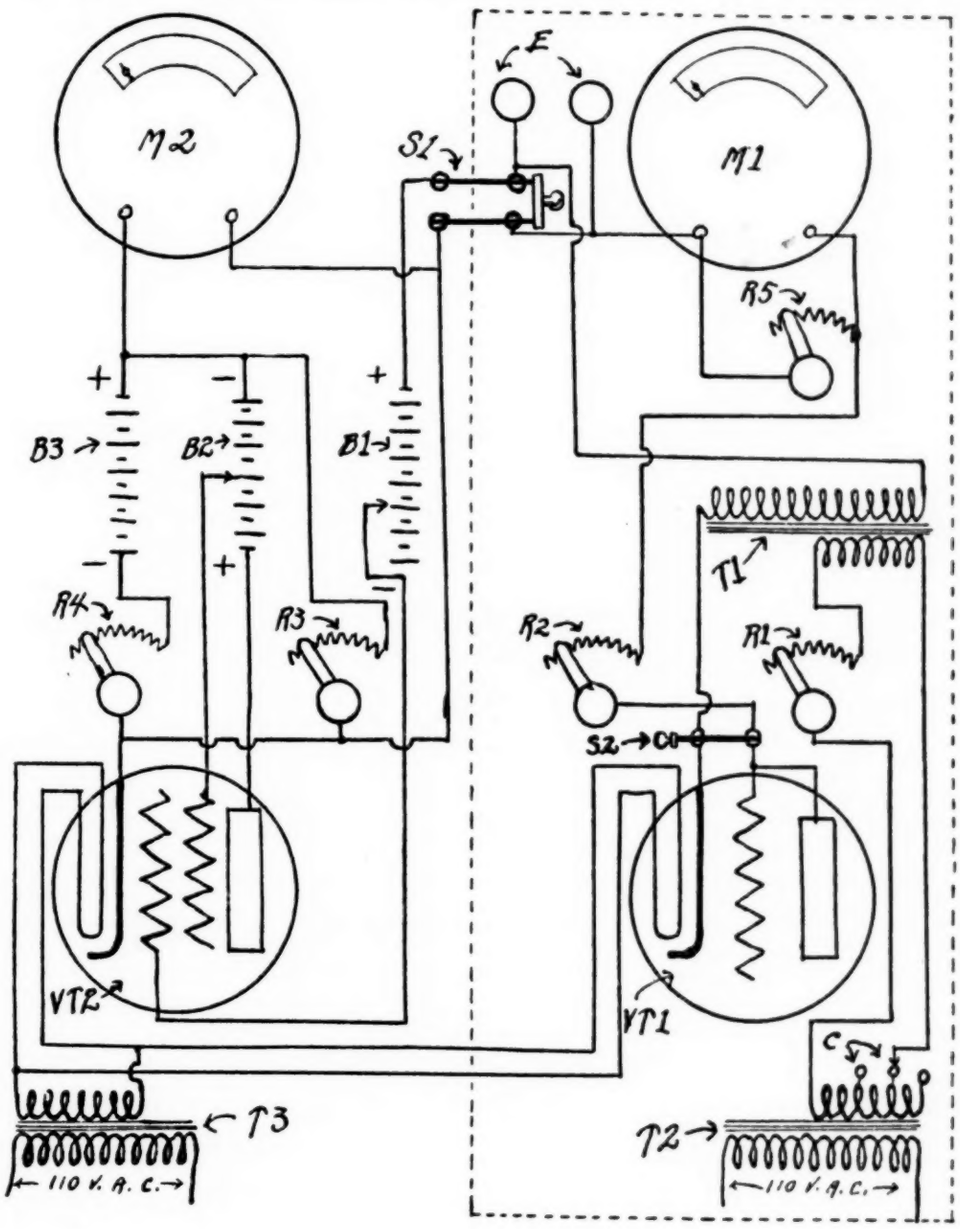


FIG. 1

vacuum-tube voltmeter, which constitutes the remaining part of the diagram, is connected in parallel across the electrodes, E. One electrode is connected to the cathode element of the screen-grid vacuum-tube, $VT2$, while the other

electrode is connected to the control grid of the tube through a *C* battery, *B*1. The potential of this battery is ordinarily about 18 volts, but should vary with the magnitude of the current passed through the tissues. A larger current means a greater EMF across the electrodes and consequently a greater *C* voltage is needed in order to maintain the efficiency of the tube. The voltage of *B*1 is not critical but should always be adjusted for maximum sensitivity.

The plate circuit of the tube, *VT*2, consists, first, of the *B* battery (*B*2, 90 v.), and second, the meter, *M*2. The battery has a mid-tap at 45 v. which is connected to the screen grid of the tube. The meter should be a sensitive galvanometer with a rotary coil which is wound with copper wire to a resistance of 1000 ohms or over.[3]

The meter, *M*2, has a variable rheostat, *R*3, connected across it for the purpose of controlling its sensitivity. When used with a 1000 ohm meter its resistance should be about 5000 ohms. A battery, *B*3, and rheostat, *R*4, are also connected in series across the meter as a means of balancing out the deflection produced by the initial plate current and adjusting the meter to zero reading. When used in relationship with a 1000 ohm meter the rheostat should have a range of resistance from 200 ohms up to 30 megohms, or thereabouts. The writer used a Pilot resistograd. The battery may be a "B" battery of about 100 volts.

*T*3 is a filament-lighting transformer having an output potential of about 2.5 volts, which is adapted for lighting the filaments of the two tubes.

In order to prepare the instrument for operation the switches, *S*1 and *S*2, are both opened, the rheostats, *R*1 and *R*4, are both set at near maximum resistance, while the rheostats, *R*3 and *R*5, are set at minimum resistance. The

[3] Increase in the resistance of this coil, up to any practicable limit when it is used in connection with a screen-grid tube, results in an increase in the sensitivity of the meter. Since the apparatus is abundantly sensitive, however, a meter of comparatively lower resistance will serve quite well. The writer rewound the coil of a Jewell galvanometer, Pattern 54, with No. 50 wire. It is suggested that one of the new pentode tubes could be substituted and used in connection with an ordinary galvanometer, having a resistance of about 50 ohms.

electrodes, *E*, are shorted, and the wire connecting to the secondary of the supply transformer, *T*2, at *C* is disconnected. The 110 v. A.C. input to *T*2 and *T*3 is now turned on and trial contacts are made of the various output potentials from *T*2, starting with the lowest (about 1 v.), until a contact is discovered which will slightly affect the meter when *R*1 is at zero resistance. *R*5 is now increased to maximum and the control rheostat, *R*1, is adjusted so that a reading of about 50 microamperes is obtained on the meter. The subject is now placed in the circuit and the current readjusted to the same value.

The apparatus is now prepared for use as an electrical stimulus, as described in the reference previously cited. Diminishing the resistance of the control rheostat, *R*1, increases the strength of the stimulus. If the maximum intensity is not sufficient it may be increased by increasing the potential obtained from *T*2. If the range of *M*1 should be insufficient it may be increased by adjusting *R*5.

In order to use the instrument as a direct-current type of psychogalvanometer the procedure described previously to the last paragraph is followed. The switch, *S*1, is closed and *R*3 is adjusted so that a small deflection of the meter, *M*2, is observed. This reading is now adjusted to zero by adjustment of the rheostat, *R*4. The rheostat, *R*3, is now increased to near maximum resistance, whereupon the psychogalvanic effects can be observed in the variations of *M*2. If the indicator hand goes beyond the scale it can be brought back by readjustment of *R*4. If the instrument is too sensitive it can be made less so by reducing the resistance of *R*3.

In order to use the instrument as an alternating-current psychogalvanometer the same procedure is followed as in the case of the D.C. instrument, except that at the beginning the variable resistance, *R*2, is set at near maximum and the switch, *S*2, is closed. This discontinues the rectification of the tube, *VT*1, and passes an alternating current through the body of the subject. The meter, *M*2, will not register alternating current and should be shorted by reducing the resis-

tance, $R5$, to zero. In order to obtain maximum sensitivity it may be necessary to readjust the voltage obtained from $B1$.

The apparatus as described is extremely sensitive, and yet it is very stable and easily controlled. The parts are inexpensive and it is not difficult to construct or learn to operate the apparatus. In comparison with other instruments the described apparatus has certain advantages that are apparently worth consideration, among which are the following:

1. The cost of construction is probably less than for any other type of psychogalvanometer.

2. The instrument is more sensitive than any other type of D.C. galvanometer.

3. The fact that the apparatus can be used as either a D.C. or A.C. galvanometer is a distinct advantage. Each type of instrument has its advantages and is especially adapted for certain purposes. There is a need for an improved and less-expensive type of D.C. galvanometer, partly because of the fact that certain of the skin phenomena seems to escape the A.C. instrument.

4. The peculiar nature of this instrument and its very wide range of control suggests the possibility of some important research in regard to the sources and constituents of the psychogalvanic phenomena.

Since the instrument is a voltage operated, rather than a current operated instrument, as all previous types have been, and since practically no current is consumed in operating the voltmeter, it is possible to study the psychogalvanic phenomena under circumstances in which different quantities of current are flowing through the tissues and when different degrees of polarization exist at the points of contact.

The writer has tentative results with the direct-current instrument which indicate that a quicker and more variable reaction is obtained when the current through the tissues is very weak. This result may possibly be due to the fact that under these circumstances there is a minimum of polarization in the skin, which fact allows the glands in the skin to play a relatively more important part. It seems probable that

two or more factors may combine to produce the typical electrical reaction. The slower reaction which is observed when the larger current is flowing through the tissues may be more largely the product of vaso-motor changes in the deeper tissues. This instrument should afford a considerable advantage in this and similar studies.

(Manuscript received June 6, 1931)

A NEW DEVICE FOR THE MEASUREMENT OF TIME INTERVALS

BY F. M. DENTON, F.C.G.I.

University of New Mexico

Pendulum methods of measuring short intervals of time are slow and tiresome and their range is limited.

The object of this article is to draw attention to an electrical method which is simple, speedy, accurate, inexpensive and of wide range. The device required is not on the market but may readily be assembled in the laboratory.

The method applies the well-known fact that when an electrical condenser of capacity C farads is charged through an ohmic resistance of R ohms by a battery of E volts the charge attains 63 percent of its final value in RC seconds—the quantity RC being known as the 'time-constant' of the circuit.

For instance, if the condenser C in Fig. 1 have a capacity of four microfarads and be charged through a resistance R

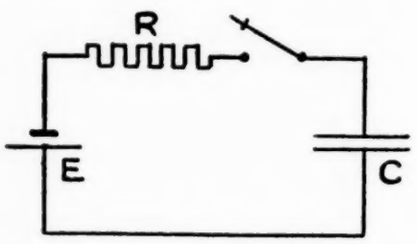


FIG. 1

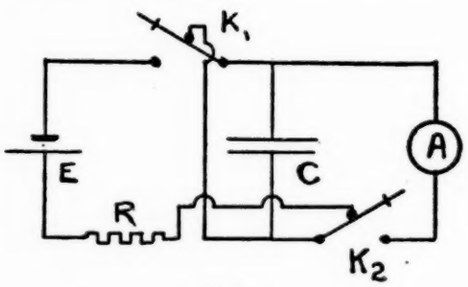


FIG. 2

of 30,000 ohms from a battery of $22\frac{1}{2}$ volts, the time required to charge the condenser to 63 percent of its final value of 9×10^{-5} coulombs will be RC ($= 30{,}000 \times 4 \times 10^{-6}$) $= 0.12$ second.

The quantity of charge that has gone into C during any time interval may be measured by allowing the condenser to discharge through a ballistic galvanometer, the magnitude of the first swing being a measure of the charge.

Instead of a ballistic galvanometer it has been found practicable to use an ordinary commercial D.C. milliammeter.

A suitable arrangement is shown in Fig. 2. Depression of the key K_1 charges the condenser. Depression of K_2 (the key K_1 being held down) allows C to discharge through the ammeter. The time interval between the depressions of K_1 and K_2 is measured directly by the throw of the ammeter needle.

The following data refer to such a device built of components usually available in a physics laboratory.

A is a miniature round pattern D.C. milliammeter of range about half a milliampere (made by the Jewell Instrument Co.). Its resistance is 15 ohms and its time of swing about one second, the damping being such that it makes some twelve swings before coming to rest on zero.

R is a resistance box or grid-leak providing fixed values from a few thousand to five million ohms.

E is a small Burgess battery of $22\frac{1}{2}$ volts. The milliammeter has an evenly divided scale of ten divisions.

The curve of Fig. 3 translates deflections into seconds for the case in which R is 30,000 ohms, the range being about 0.2 second. The time scale may be multiplied by any desired constant k by making R equal to 30,000 k. For instance, with R equal to 30,000 ohms a throw of 5.5 divisions means a time interval of 0.1 second, while with R 3000 ohms the same deflection means 0.01 second; with R 300,000 it means one second and with R 3 megohms, ten seconds.

Experiment shows that if an accuracy of five percent in the middle of the scale is sufficient one and the same calibration curve may be used for a wide range of values of R. For great accuracy special curves must be drawn for values of R which are widely different, especially when R is of the same order of magnitude as the insulation resistance of the condenser itself.

It is important to work on the lower part of the curve of charge well below the knee. This condition is insured

by proper choice of the capacity C and the voltage E. Thoroughly reliable paper condensers should be used.

The time intervals used in plotting the curve of Fig. 3

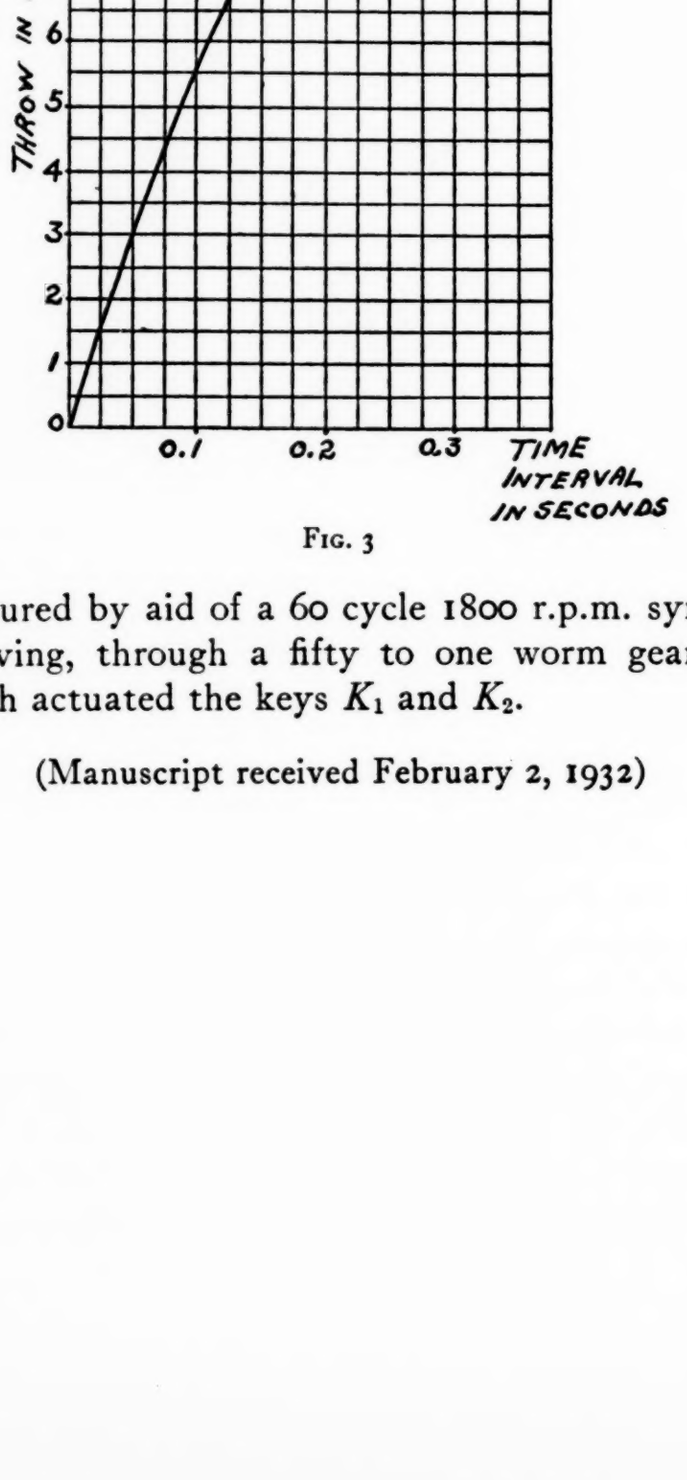


FIG. 3

were measured by aid of a 60 cycle 1800 r.p.m. synchronous motor driving, through a fifty to one worm gear, a turntable which actuated the keys K_1 and K_2.

(Manuscript received February 2, 1932)